

!!SEQUENCE LIST 1.0 !!FINDPATTERNS on Swiss-Prot: * allowing 0 mismatches ! <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C,T,S)(R,K,H											
1	1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C,T,S)(R,K,H										
SW:ANKG_PIG	ck: 8859	len: 133	finds: 1	! Q09138	sus scrofa (pig). 5'-an	SW:CHRR_RHOSH	ck: 8138	len: 213	finds: 1	! P40685	rhodobacter sphaerol
SW:ACP_BACSU	ck: 8148	len: 77	finds: 1	! P80643	bacillus subtilis. acyl	SW:COAT_TCV	ck: 1818	len: 351	finds: 1	! P06663	turnip crinkle virus
SW:ADH2_DROMO	ck: 8528	len: 253	finds: 1	! P09369	drosoophila mojavensis	SW:COBT_ECOLI	ck: 7029	len: 359	finds: 1	! P36562	escherichia coli. n1
SW:ALR_TREPA	ck: 255	len: 357	finds: 1	! Q56346	treponema pallidum. aia	SW:COX3_SCHCO	ck: 7659	len: 268	finds: 1	! P14058	schizophyllum commun
SW:AMIA_SALTY	ck: 5663	len: 289	finds: 1	! P33772	salmonella typhimurium.	SW:COX4_SCHPO	ck: 7810	len: 164	finds: 1	! P79010	schizosaccharomyces
SW:AMPC_SERMA	ck: 107	len: 376	finds: 1	! P18539	serratia marcescens. be	SW:CPCL_CANPG	ck: 1605	len: 114	finds: 1	! P81580	cancer pagurus (rock
SW:AMPD_CITFR	ck: 4831	len: 187	finds: 1	! P00831	citrobacter freundli. a	SW:CPTN_PEA	ck: 5889	len: 402	finds: 1	! P21727	pisum sativum (garde
SW:ANX2_XENLA	ck: 6452	len: 339	finds: 1	! P24801	xenopus laevis (african	SW:CSCR_ECOLI	ck: 8969	len: 331	finds: 1	! P40715	escherichia coli. su
SW:ARG1_XENLA	ck: 4321	len: 360	finds: 1	! Q91553	xenopus laevis (african	SW:CTAA_BACFI	ck: 716	len: 297	finds: 1	! Q04443	bacillus firmus. cyt
SW:ARG2_XENLA	ck: 4641	len: 360	finds: 1	! Q91554	xenopus laevis (african	SW:CTAA_BACSU	ck: 1522	len: 306	finds: 1	! P12946	bacillus subtilis. c
SW:ARG3_XENLA	ck: 4724	len: 360	finds: 1	! Q91555	xenopus laevis (african	SW:CU04_BIACR	ck: 1926	len: 127	finds: 1	! P80675	blaberus crantlifer.
SW:ARGC_ECOLI	ck: 6342	len: 334	finds: 1	! P11446	escherichia coli. n-ace	SW:CYR2_RHOYE	ck: 2620	len: 92	finds: 1	! P00098	rhodocyclus tenuis (
SW:AT93_HUMAN	ck: 2178	len: 142	finds: 1	! P48201	homo sapiens (human). a	SW:CYNR_ECOLI	ck: 4304	len: 299	finds: 1	! P27111	escherichia coli. cy
SW:ATP8_PODAN	ck: 8763	len: 50	finds: 1	! Q02653	podospora anserina. atp	SW:DACX_STRSK	ck: 4817	len: 291	finds: 1	! P39042	streptomyces sp. (st
SW:ATP8_PONPA	ck: 3336	len: 68	finds: 1	! P92694	pongo pygmaeus abelli	SW:DBP5_YEAST	ck: 2958	len: 482	finds: 1	! P20449	saccharomyces cerevi
SW:ATP8_PONPY	ck: 3460	len: 68	finds: 1	! P92896	pongo pygmaeus pygmaeus	SW:DFP_STMXU	ck: 9135	len: 145	finds: 1	! Q54433	streptococcus mutans
SW:ATP8_PONPY	ck: 3550	len: 68	finds: 1	! Q35584	pongo pygmaeus (orangut	SW:DHAT_CITFR	ck: 101	len: 387	finds: 1	! P45513	citrobacter freundli
SW:ATPD_HUMAN	ck: 1686	len: 168	finds: 1	! P30049	homo sapiens (human). a	SW:DHAT_KLEPN	ck: 846	len: 387	finds: 1	! Q59477	klebsiella pneumonia
SW:ATP2_BACP3	ck: 4392	len: 127	finds: 1	! P09354	bacillus ps3 (thermophil	SW:DHBB_HUMAN	ck: 3434	len: 310	finds: 1	! P37058	homo sapiens (human)
SW:BMPA_XENLA	ck: 9714	len: 398	finds: 1	! P25703	xenopus laevis (african	SW:DHBB_MOUSE	ck: 9317	len: 305	finds: 1	! P70385	mus musculus (mouse)
SW:BMPB_XENLA	ck: 1620	len: 398	finds: 1	! P30884	xenopus laevis (african	SW:DHBB3_RAT	ck: 3366	len: 306	finds: 1	! Q54939	rattus norvegicus (r
SW:BPHB_PSEPS	ck: 1527	len: 275	finds: 1	! P08694	pseudomonas pseudoalcali	SW:DHX_MOUSE	ck: 4788	len: 312	finds: 1	! Q70503	mus musculus (mouse)
SW:BR3A_HUMAN	ck: 2610	len: 423	finds: 1	! Q01851	homo sapiens (human). h	SW:DHX3_MOUSE	ck: 9317	len: 305	finds: 1	! P70385	mus musculus (mouse)
SW:BRAL_BRAFL	ck: 7365	len: 448	finds: 1	! Q17134	branchiostoma floridae	SW:DHCA_HUMAN	ck: 3202	len: 276	finds: 1	! P16152	homo sapiens carbohy
SW:BRAL_BRAFL	ck: 391	len: 440	finds: 1	! P80492	branchiostoma floridae	SW:DHCA_MOUSE	ck: 204	len: 276	finds: 1	! P48758	mus musculus (mouse)
SW:BSR_BACCE	ck: 8425	len: 140	finds: 1	! P33967	bacillus cereus. blasti	SW:DHCA_RABIT	ck: 1426	len: 276	finds: 1	! P47844	oryctolagus cuniculu
SW:CADI_DICDI	ck: 3813	len: 213	finds: 1	! P54657	dictyostelium discoideu	SW:DHCA_RAT	ck: 3202	len: 276	finds: 1	! P47727	rattus norvegicus (r
SW:CAMG_MOUSE	ck: 283	len: 294	finds: 1	! P49070	mus musculus (mouse). c	SW:DHCA_RAT	ck: 3202	len: 276	finds: 1	! P47727	rattus norvegicus (r
SW:CDD_BACSU	ck: 9048	len: 136	finds: 1	! P19079	bacillus subtilis. cyt	SW:DHK1_STRVN	ck: 3104	len: 272	finds: 1	! P16542	streptomyces violace
SW:CDD_MYCPI	ck: 6718	len: 133	finds: 1	! P47718	mycoplasma pirum. cyt	SW:DHK1_STRVN	ck: 3104	len: 272	finds: 1	! P16542	streptomyces violace
SW:CDSA_HAEIN	ck: 8973	len: 288	finds: 1	! P44937	haemophilus influenzae.	SW:DHKR_STGCM	ck: 4473	len: 261	finds: 1	! P41177	streptomyces cinamo
SW:CDSN_PIG	ck: 9217	len: 225	finds: 1	! Q19084	sus scrofa (pig). corne	SW:DITE_BACSU	ck: 2631	len: 252	finds: 1	! P36577	bacillus subtilis. d
SW:CG21_SCHPO	ck: 7358	len: 415	finds: 1	! P24865	schizosaccharomyces pom	SW:DSBB_VIBAL	ck: 2736	len: 165	finds: 1	! Q56578	vibrio alginolyticus
SW:CH19_DROGR	ck: 4664	len: 196	finds: 1	! P13427	drosophila grimshawi (f	SW:DUS1_HUMAN	ck: 5401	len: 367	finds: 1	! P28562	homo sapiens (human)
						SW:DUS1_MOUSE	ck: 5871	len: 367	finds: 1	! P28563	mus musculus (mouse)
						SW:DUS1_RAT	ck: 6443	len: 367	finds: 1	! Q64623	rattus norvegicus (r

SW:DVR1_BRARE	ck: 2951	len: 355	finds: 1	P35621	brachydanio rerio (zebr)
SW:EL128_DROME	ck: 68	len: 255	finds: 1	P08761	drosophila melanogaster
SW:ELAC_TRIYU	ck: 5524	len: 102	finds: 1	P02143	trichosturus vulpecula (c
SW:ENP2_YEAST	ck: 5386	len: 215	finds: 1	P39704	saccharomyces cerevisiae
SW:ES1_BRARE	ck: 1627	len: 270	finds: 1	P090257	brachydanio rerio (zebr)
SW:EFPA_MYCLE	ck: 3764	len: 318	finds: 1	P035096	mycobacterium leprae. e
SW:EXBD_NEICO	ck: 4894	len: 144	finds: 1	P006434	neisseria gonorrhoeae.
SW:EXBD_NEIME	ck: 5831	len: 144	finds: 1	P95376	neisseria meningitidis.
SW:FABG_ECOLI	ck: 3492	len: 244	finds: 1	P25716	escherichia coli. 3-oxo
SW:FABG_HAEIN	ck: 4159	len: 242	finds: 1	P43713	haemophilus influenzae.
SW:FABG_PSEAE	ck: 9922	len: 247	finds: 1	P054438	pseudomonas aeruginosa.
SW:FABG_VIBHA	ck: 5984	len: 244	finds: 1	P55336	vibrio harveyi. 3-oxoac
SW:FENR_ECOLI	ck: 3071	len: 247	finds: 1	P28861	escherichia coli. ferre
SW:FENR_SHIFL	ck: 1251	len: 135	finds: 1	P28901	shigella flexneri. feri
SW:FIBR_AGKCO	ck: 7566	len: 203	finds: 1	P28891	agkistrodon contortrix
SW:FIMH_SALTY	ck: 6074	len: 335	finds: 1	P37925	salmonella typhimurium.
SW:FIH_BACSV	ck: 1150	len: 208	finds: 1	P23449	bacillus subtilis. prob
SW:FOS_AVINK	ck: 9324	len: 322	finds: 1	P23050	avian retrovirus nk24.
SW:FOS_CHICK	ck: 7007	len: 367	finds: 1	P11939	gallus gallus (chicken)
SW:FOS_HUMAN	ck: 4677	len: 380	finds: 1	P01100	homo sapiens (human). F
SW:FRRC_MAIRE	ck: 4642	len: 152	finds: 1	P41347	zea mays (maize). ferre
SW:GF88_BOVIN	ck: 9403	len: 375	finds: 1	P01836	bos taurus (bovine). gr
SW:GF88_HUMAN	ck: 1814	len: 375	finds: 1	P014793	homo sapiens (human). g
SW:GF88_MOUSE	ck: 2293	len: 376	finds: 1	P008689	mus musculus (mouse). g
SW:GF88_PAPHA	ck: 1463	len: 375	finds: 1	P018828	papio hamadryas (hamad
SW:GF88_PIG	ck: 1805	len: 375	finds: 1	P018831	sus scrofa (pig). growt
SW:GF88_RAT	ck: 1496	len: 376	finds: 1	P035312	rattus norvegicus (rat)
SW:GF88_SHEEP	ck: 1548	len: 375	finds: 1	P018830	ovis aries (sheep). grc
SW:GDPB_PANTR	ck: 149	len: 123	finds: 1	P028914	pan troglodytes (chimp)
SW:GDDA_ECOLI	ck: 4443	len: 339	finds: 1	P37606	escherichia coli. glyce
SW:GPRD_HUMAN	ck: 2863	len: 355	finds: 1	P49238	homo sapiens (human). f
SW:GPRD_RAT	ck: 4478	len: 354	finds: 1	P33411	rattus norvegicus (rat)
SW:GRAB_MOUSE	ck: 92	len: 247	finds: 1	P04187	mus musculus (mouse). g
SW:HDHA_CLOSO	ck: 1047	len: 267	finds: 1	P50200	clostridium sordeifili.
SW:HS2A_HUMAN	ck: 9949	len: 296	finds: 1	P04233	homo sapiens (human). h
SW:HS8_SALTY	ck: 516	len: 359	finds: 1	P10369	salmonella typhimurium.
SW:HSO_ECOLI	ck: 7651	len: 228	finds: 1	P52094	escherichia coli. histid
SW:HMPH_HUMAN	ck: 3785	len: 270	finds: 1	P003014	homo sapiens (human)
SW:HMPH_MOUSE	ck: 8616	len: 271	finds: 1	P43120	mus musculus (mouse)
SW:HPRK_BACSV	ck: 4393	len: 309	finds: 1	P034483	bacillus subtilis. h
SW:HPRK_RHILV	ck: 1548	len: 370	finds: 1	P28153	rhizobium leguminosa
SW:HYCD_ECOLI	ck: 675	len: 307	finds: 1	P16430	escherichia coli. fo
SW:IAA1_WHEAT	ck: 3320	len: 124	finds: 1	P01085	triticum aestivum (w
SW:IAA2_WHEAT	ck: 4825	len: 123	finds: 1	P01083	triticum aestivum (w
SW:IAA5_WHEAT	ck: 4073	len: 124	finds: 1	P01084	triticum aestivum (w
SW:ICE2_RAT	ck: 8827	len: 312	finds: 1	P55215	rattus norvegicus (r
SW:ICW3_PSORE	ck: 7697	len: 207	finds: 1	P10822	psophocarpus tetrago
SW:IDI_MYCTU	ck: 9024	len: 203	finds: 1	P72002	mycobacterium tuberc
SW:IDNR_ECOLI	ck: 2843	len: 332	finds: 1	P39343	escherichia coli. 1-
SW:IMAN_ECOLI	ck: 6528	len: 131	finds: 1	P09182	escherichia coli. co
SW:IMPB_SALTY	ck: 6434	len: 424	finds: 1	P18642	salmonella typhimuri
SW:IPRA_SAGSA	ck: 1378	len: 181	finds: 1	P31608	sagittaria sagittifo
SW:IS12_STRAL	ck: 4079	len: 256	finds: 1	P23393	streptomyces albus g
SW:ISRA_SYNY3	ck: 4153	len: 342	finds: 1	P052274	synecocystis sp. (s
SW:KDD_BACSV	ck: 3484	len: 254	finds: 1	P50842	bacillus subtilis. 2
SW:KRL_HSVS	ck: 5295	len: 467	finds: 1	P008097	bovine herpesvirus t
SW:LSRA_MYCTU	ck: 2319	len: 202	finds: 1	P010764	mycobacterium tuberc
SW:MAUF_PARBE	ck: 7474	len: 277	finds: 1	P28897	paracoccus denitrifi
SW:MDH_RHILV	ck: 1419	len: 320	finds: 1	P03525	rhizobium leguminosa
SW:MESJ_HAEIN	ck: 1410	len: 430	finds: 1	P44689	haemophilus influenz
SW:MIPA_MYXXA	ck: 8809	len: 236	finds: 1	P38371	myxococcus xanthus.
SW:MML6_MYCTU	ck: 4418	len: 397	finds: 1	P010773	mycobacterium tuberc
SW:MMDA_METER	ck: 9524	len: 288	finds: 1	P55818	methyllobacterium ext
SW:NEUB_XENIA	ck: 1414	len: 120	finds: 1	P43443	xenopus laevis (afri
SW:NIRY_KLEPN	ck: 8799	len: 229	finds: 1	P09135	klebsiella pneumonia
SW:NOB1_ASPPA	ck: 7169	len: 271	finds: 1	P000278	aspergillus parasiti
SW:NOUM_COTJA	ck: 6471	len: 72	finds: 1	P24968	coturnix coturnix ja.
SW:NOUM_ANSE	ck: 1655	len: 214	finds: 1	P03651	anser caerulescens (
SW:NOUM_APIPI	ck: 6545	len: 167	finds: 1	P34857	apis mellifera ligus
SW:NOUM_NEUCR	ck: 3961	len: 263	finds: 1	P40915	neurospora crassa. n
SW:NOUM_BRAOL	ck: 9647	len: 215	finds: 1	P42027	brassica oleracea (c
SW:NOUJ_ECOLI	ck: 2909	len: 184	finds: 1	P33605	escherichia coli. na
SW:NOZM_NEUCR	ck: 7750	len: 201	finds: 1	P19968	neurospora crassa. n

SW:OPPB_ECOLI	ck: 5147	len: 306	finds: 1	P31132	escherichia coli. oligo
SW:OPPB_SALTY	ck: 4920	len: 306	finds: 1	P08005	salmonella typhimurium.
SW:OPSL_CALVI	ck: 9791	len: 371	finds: 1	P22269	calliphora vicina (blue
SW:OPSL_DROME	ck: 4358	len: 373	finds: 1	P06002	drosoophila melanogaster
SW:OPSL_DROPS	ck: 7241	len: 374	finds: 1	P28678	drosoophila pseudoobscur
SW:OTCA_VIBS2	ck: 7498	len: 301	finds: 1	P96172	vibrio sp. (strain 2693
SW:OTC_SCHPO	ck: 9240	len: 327	finds: 1	P31317	schizosaccharomyces pom
SW:OXND_BOVIN	ck: 9943	len: 341	finds: 1	P31228	bos taurus (bovine). d-
SW:OXND_HUMAN	ck: 7167	len: 341	finds: 1	Q09489	homo sapiens (human). d-
SW:P2Y6_HUMAN	ck: 7545	len: 328	finds: 1	Q15077	homo sapiens (human). F
SW:PETP_RHOCA	ck: 1875	len: 166	finds: 1	P31078	rhodobacter capsulatus
SW:PEX3_YEAST	ck: 166	len: 441	finds: 1	P28795	saccharomyces cerevisiae
SW:PHBB_ALCEU	ck: 2829	len: 246	finds: 1	P14697	alcaligenes eutrophus.
SW:PHNK_ECOLI	ck: 804	len: 252	finds: 1	P16678	escherichia coli. phosp
SW:PI7A_ORYSA	ck: 1305	len: 263	finds: 1	Q40708	oryza sativa (rice). ph
SW:PLAS_SYNY3	ck: 6628	len: 126	finds: 1	P21697	synechocystis sp. (str
SW:PLEK_HUMAN	ck: 1536	len: 350	finds: 1	P08567	homo sapiens (human). F
SW:PLYA_MYCPO	ck: 8743	len: 299	finds: 1	Q12639	mycosphaerella pinodes
SW:PME_ASPAC	ck: 2447	len: 331	finds: 1	Q12535	aspergillus aculeatus.
SW:PME_ASPTU	ck: 609	len: 331	finds: 1	P17872	aspergillus tubingensis
SW:PSPB_BOVIN	ck: 7861	len: 79	finds: 1	P15781	bos taurus (bovine). pu
SW:PTB_BACSV	ck: 158	len: 299	finds: 1	P54530	bacillus subtilis. prob
SW:PTI6_HUMAN	ck: 300	len: 376	finds: 1	P35237	homo sapiens (human). P
SW:PTI8_HUMAN	ck: 4145	len: 374	finds: 1	P50452	homo sapiens (human). C
SW:PURT_METTH	ck: 5550	len: 248	finds: 1	Q26272	methanobacterium thermo
SW:PYRB_ECOLI	ck: 3945	len: 310	finds: 1	P00479	escherichia coli. aspar
SW:PYRB_MYCTU	ck: 1547	len: 319	finds: 1	P71808	mycobacterium tuberculo
SW:PYRB_PYRAB	ck: 9304	len: 308	finds: 1	P77918	pyrococcus abyssi. aspa
SW:PYRB_PYRHO	ck: 9980	len: 308	finds: 1	O58451	pyrococcus horikoshii.
SW:PYRB_SALTY	ck: 2583	len: 310	finds: 1	P08420	salmonella typhimurium.
SW:PYRB_SERMA	ck: 4826	len: 305	finds: 1	P19910	serattia marcescens. as
SW:PYRB_VIBS2	ck: 5096	len: 310	finds: 1	P96174	vibrio sp. (strain 2693
SW:PYRB2_BACSV	ck: 2658	len: 256	finds: 1	P25983	bacillus subtilis. dthy
SW:RAS1_DROME	ck: 2533	len: 336	finds: 1	Q27297	drosoophila melanogaster
SW:RBGR_CHRYI	ck: 4295	len: 302	finds: 1	P25544	chromatium vinosum. rut
SW:RBSD_SOLTU	ck: 4699	len: 181	finds: 1	P10647	solanum tuberosum (pot
SW:RBS1_LYCES	ck: 5240	len: 181	finds: 1	P08706	lycopersicon esculentum
SW:RBS1_PETSP	ck: 2967	len: 180	finds: 1	P04714	petunia sp. (petunia
SW:RBS1_SOLTU	ck: 4863	len: 181	finds: 1	P26574	solanum tuberosum (p
SW:RBS2_LYCES	ck: 4535	len: 180	finds: 1	P07179	lycopersicon esculen
SW:RBS3_SOLTU	ck: 4433	len: 181	finds: 1	P22764	solanum tuberosum (p
SW:RBSK_SCHPO	ck: 3214	len: 318	finds: 1	Q60116	schizosaccharomyces
SW:REB3_MUSAC	ck: 2518	len: 180	finds: 1	Q24045	musa acuminata (bana
SW:REB5_DROME	ck: 7903	len: 298	finds: 1	Q94913	drosoophila melanogas
SW:REHY_TORRU	ck: 4850	len: 218	finds: 1	P52574	tortula ruralis (mos
SW:REP2_YEAST	ck: 6596	len: 296	finds: 1	P03872	saccharomyces cerevi
SW:REPB_NEIGO	ck: 5086	len: 346	finds: 1	P37761	neisseria gonorrhoea
SW:RHL_MACMU	ck: 4371	len: 416	finds: 1	Q28849	macaca mulatta (rhes
SW:RH07_HUMAN	ck: 2901	len: 227	finds: 1	P52198	homo sapiens (human)
SW:RIE7_ARCTU	ck: 4460	len: 219	finds: 1	Q28272	archaeoglobus fulgid
SW:RIPA_LUFXY	ck: 429	len: 277	finds: 1	Q00465	luffa cylindrica (em
SW:RIPB_LUFXY	ck: 7681	len: 250	finds: 1	P22851	luffa cylindrica (em
SW:RR21_SPIOL	ck: 718	len: 256	finds: 1	P24613	splnacia oleracea (s
SW:RR8_MARPO	ck: 6002	len: 132	finds: 1	P06362	marichanta polymorph
SW:RS13_SCHPO	ck: 3858	len: 150	finds: 1	P28189	schizosaccharomyces
SW:RS3_ACHAX	ck: 7310	len: 257	finds: 1	P41117	acholoplasma axanthu
SW:RS8_SYNY3	ck: 8926	len: 133	finds: 1	P73307	synechocystis sp. (s
SW:RVB_TRETH	ck: 3575	len: 324	finds: 1	Q56214	thermus aquaticus (s
SW:S27A_ECOLI	ck: 54	len: 217	finds: 1	P26428	escherichia coli. sl
SW:SECY_RICPR	ck: 3620	len: 433	finds: 1	Q9205	rickettsia prowazeki
SW:SIXA_ECOLI	ck: 9360	len: 161	finds: 1	P76502	escherichia coli. ph
SW:SIXA_HAEIN	ck: 2974	len: 164	finds: 1	P44164	haemophilus influenza
SW:SPAL_SALTY	ck: 7477	len: 432	finds: 1	P39444	salmonella typhimuri
SW:SPIR_SPIIME	ck: 9527	len: 241	finds: 1	P21625	spiroplasma mellifer
SW:SRP_CHLPS	ck: 9648	len: 134	finds: 1	P28164	chlamydia psittaci.
SW:SSAU_SALTY	ck: 3263	len: 352	finds: 1	P96069	salmonella typhimuri
SW:SSRI_HUMAN	ck: 8652	len: 391	finds: 1	P30872	homo sapiens (human)
SW:SSRI_MOUSE	ck: 8110	len: 391	finds: 1	P30873	mus musculus (mouse)
SW:SSRI_RAT	ck: 9240	len: 391	finds: 1	P28646	rattus norvegicus (r
SW:SSRB_CANFA	ck: 8392	len: 183	finds: 1	P23438	canis familiaris (do
SW:SSRB_HUMAN	ck: 8093	len: 183	finds: 1	P43308	homo sapiens (human)
SW:STI4_SOLTU	ck: 5285	len: 214	finds: 1	Q41495	solanum tuberosum (p
SW:SUPL_HAEIN	ck: 4	len: 311	finds: 1	P44847	haemophilus influenza

SW:SUH3_RAT	CK: 800	len: 285	finds: 1	! P50235 rattus norvegicus (rat)
SW:TRC1_CHV13	CK: 7233	len: 278	finds: 1	! P52283 chlorella virus il-3a (c
SW:TCF1_HUMAN	CK: 5007	len: 269	finds: 1	! P36402 homo sapiens (human). t
SW:TELT_HUMAN	CK: 3375	len: 167	finds: 1	! O15273 homo sapiens (human). t
SW:TELT_MOUSE	CK: 5613	len: 167	finds: 1	! O70548 mus musculus (mouse). t
SW:TRAM_AGR6	CK: 4597	len: 102	finds: 1	! Q57471 agrobacterium tumefacie
SW:TRND_SYNY3	CK: 6210	len: 231	finds: 1	! P72828 synechocystis sp. (stra
SW:TRND_SCHPO	CK: 7997	len: 415	finds: 1	! O13947 schizosaccharomyces pom
SW:TRND_YEAST	CK: 8415	len: 417	finds: 1	! O12093 saccharomyces cerevisia
SW:TRNH_DATST	CK: 5293	len: 268	finds: 1	! P50165 datura stramonium (jims
SW:TRPF_CRYNE	CK: 2342	len: 312	finds: 1	! P27710 cryptococcus neoformans
SW:TRFL_CANFA	CK: 4988	len: 371	finds: 1	! P43698 canis familiaris (dog). t
SW:TRFL_HUMAN	CK: 5439	len: 371	finds: 1	! P43699 homo sapiens (human). t
SW:TRFL_MOUSE	CK: 7622	len: 372	finds: 1	! P50220 mus musculus (mouse). t
SW:TRFL_RAT	CK: 6608	len: 372	finds: 1	! P23441 rattus norvegicus (rat). t
SW:TXF8_DENAN	CK: 8703	len: 81	finds: 1	! P01404 dendroaspis angusticeps
SW:UBC7_DROME	CK: 1891	len: 153	finds: 1	! P52487 drosophila melanogaster
SW:UMUD_ECOLI	CK: 8403	len: 159	finds: 1	! P04153 escherichia coli. umud
SW:USO2_HSVK	CK: 5996	len: 303	finds: 1	! P32517 equine herpesvirus type
SW:VANS_ENTFC	CK: 3601	len: 384	finds: 1	! Q08240 enterococcus faecium (s
SW:VE2_HPV03	CK: 4819	len: 383	finds: 1	! P36778 human papillomavirus ty
SW:VE2_HPV10	CK: 273	len: 376	finds: 1	! P36781 human papillomavirus ty
SW:VE2_HPV18	CK: 5006	len: 365	finds: 1	! P06790 human papillomavirus ty
SW:VE2_HPV28	CK: 1816	len: 376	finds: 1	! P25482 human papillomavirus ty
SW:VE2_HPV31	CK: 9656	len: 372	finds: 1	! P17383 human papillomavirus ty
SW:VE2_HPV45	CK: 5828	len: 368	finds: 1	! P36794 human papillomavirus ty
SW:VE2_HPV51	CK: 1816	len: 358	finds: 1	! P26547 human papillomavirus ty
SW:VE2_HPV52	CK: 4491	len: 368	finds: 1	! P36796 human papillomavirus ty
SW:VEA2_BPP22	CK: 7758	len: 317	finds: 1	! Q03544 bacteriophage p22. eaa
SW:VGG_BP64	CK: 4484	len: 177	finds: 1	! P03644 bacteriophage g4. major
SW:VGL1_HSV2H	CK: 501	len: 372	finds: 1	! P13291 herpes simplex virus (t
SW:VGLM_HSVSA	CK: 7071	len: 366	finds: 1	! Q01017 herpesvirus salmtili (st
SW:VIB8_AGR75	CK: 5994	len: 237	finds: 1	! P17798 agrobacterium tumefacie
SW:VIB8_AGR76	CK: 2646	len: 230	finds: 1	! P09781 agrobacterium tumefacie
SW:VIB8_AGR79	CK: 2826	len: 230	finds: 1	! P03357 agrobacterium tumefacie
SW:VID1_AGRRA	CK: 2761	len: 147	finds: 1	! P13461 agrobacterium rhizogene
SW:VID1_AGR75	CK: 2312	len: 147	finds: 1	! P18591 agrobacterium tumefa
SW:VID1_AGR76	CK: 4145	len: 147	finds: 1	! P06667 agrobacterium tumefa
SW:VLF1_NPROP	CK: 889	len: 374	finds: 1	! O10330 oryza pseudotsugata
SW:VMA2_BRSVA	CK: 6337	len: 186	finds: 1	! P29792 bovine respiratory s
SW:VMA2_HRSVA	CK: 9828	len: 194	finds: 1	! P04545 human respiratory sy
SW:VMA2_ORBVM	CK: 6559	len: 186	finds: 1	! Q84132 ovine respiratory sy
SW:VMEM_CVB	CK: 566	len: 106	finds: 1	! P3789 chrysanthemum virus
SW:VMEM_LVX	CK: 8138	len: 108	finds: 1	! P27331 lily virus x. hypoch
SW:VMEM_PVWR	CK: 3415	len: 109	finds: 1	! P17327 potato virus m (stra
SW:VP8_BPPH6	CK: 6252	len: 149	finds: 1	! P07579 bacteriophage phi-6.
SW:VPRR_SALDU	CK: 7061	len: 297	finds: 1	! P24417 salmonella dublin. v
SW:VPRR_SALTY	CK: 7032	len: 297	finds: 1	! P13041 salmonella typhimuri
SW:WN11_CHICK	CK: 4630	len: 354	finds: 1	! P49339 gallus gallus (chick
SW:WN11_COTJA	CK: 4030	len: 354	finds: 1	! P51891 coturnix coturnix ja
SW:WN11_HUMAN	CK: 3073	len: 354	finds: 1	! O96014 homo sapiens (human)
SW:WN11_MOUSE	CK: 1689	len: 354	finds: 1	! P48615 mus musculus (mouse)
SW:WN14_DROME	CK: 8394	len: 389	finds: 1	! P40589 drosophila melanogas
SW:Y021_NPROP	CK: 562	len: 298	finds: 1	! O10280 oryza pseudotsugata
SW:Y040_HUMAN	CK: 4815	len: 153	finds: 1	! O15053 homo sapiens (human)
SW:Y209_MYCGE	CK: 590	len: 308	finds: 1	! P47451 mycoplasma genitaliu
SW:Y209_MYCPN	CK: 6533	len: 309	finds: 1	! P75485 mycoplasma pneumonia
SW:Y280_MYCGE	CK: 9511	len: 265	finds: 1	! P47522 mycoplasma genitaliu
SW:Y373_TREPA	CK: 4876	len: 477	finds: 1	! O83388 treponema pallidum.
SW:Y425_MYCPN	CK: 2944	len: 450	finds: 1	! P75172 mycoplasma pneumonia
SW:Y464_MYCPN	CK: 2836	len: 385	finds: 1	! P47702 mycoplasma genitaliu
SW:Y464_MYCPN	CK: 5866	len: 385	finds: 1	! P75112 mycoplasma pneumonia
SW:Y4EK_RHISN	CK: 5477	len: 248	finds: 1	! P55434 rhizobium sp. (strai
SW:Y4TG_RHISN	CK: 6923	len: 231	finds: 1	! P55661 rhizobium sp. (strai
SW:Y503_MERJA	CK: 6767	len: 406	finds: 1	! Q57926 methanococcus jannas
SW:Y70A_MERJA	CK: 4119	len: 102	finds: 1	! P81311 methanococcus jannas
SW:Y788_BORBU	CK: 1619	len: 440	finds: 1	! O51728 borrelia burgdorferi
SW:Y79B_MERJA	CK: 2865	len: 170	finds: 1	! P81333 methanococcus jannas
SW:Y953_MYCTU	CK: 883	len: 282	finds: 1	! P71557 mycobacterium tuberc
SW:Y988_MERJA	CK: 9981	len: 329	finds: 1	! Q58395 methanococcus jannas
SW:Y932_PSEAE	CK: 6657	len: 122	finds: 1	! P21485 pseudomonas aerugino
SW:YABP_ECOLI	CK: 98	len: 216	finds: 1	! P39220 escherichia coli. hy

SW:YBCO_RHOCA	ck: 2127	len: 192	finds: 1	P26166 rhodobacter capsulatus
SW:YBFT_BACSU	ck: 3388	len: 249	finds: 1	O31458 bacillus subtilis. hypc
SW:YCO8_XLEPN	ck: 4280	len: 373	finds: 1	Q48454 klebsiella pneumoniae.
SW:YCL1_MEUJA	ck: 2920	len: 299	finds: 1	Q58618 methanococcus jannaschii
SW:YCB0_PORPU	ck: 4681	len: 291	finds: 1	P51321 porphyra purpurea. hypc
SW:YCBQ_ECOLI	ck: 3224	len: 182	finds: 1	P75855 escherichia coli. hypot
SW:YCGZ_ECOLI	ck: 1943	len: 78	finds: 1	P75991 escherichia coli. hypot
SW:YCU1_ECOLI	ck: 6021	len: 262	finds: 1	P51983 escherichia coli. hypot
SW:YCUZ_ECOLI	ck: 51	len: 299	finds: 1	P77333 escherichia coli. hypot
SW:YCO5_YEAST	ck: 1969	len: 317	finds: 1	P25616 saccharomyces cerevisiae
SW:YDM1_SCHPO	ck: 6054	len: 131	finds: 1	P87132 schizosaccharomyces pom
SW:YDSA_SCHPO	ck: 8869	len: 174	finds: 1	O14185 schizosaccharomyces pom
SW:YDTG_SCHPO	ck: 3748	len: 344	finds: 1	O14220 schizosaccharomyces pom
SW:YER4_YEAST	ck: 5788	len: 128	finds: 1	P40057 saccharomyces cerevisiae
SW:YF58_MYCTU	ck: 497	len: 148	finds: 1	Q10772 mycobacterium tuberculo
SW:YG22_BPSPI	ck: 7782	len: 148	finds: 1	P31653 bacteriophage sp01. hyp
SW:YKG9_YEAST	ck: 6897	len: 107	finds: 1	P53138 saccharomyces cerevisiae
SW:YGLR_STEKO	ck: 9954	len: 66	finds: 1	Q05954 streptomyces coelicolor
SW:YHAI_CRYPA	ck: 3269	len: 319	finds: 1	P10941 cryptonectria parasitid
SW:YHCT_BACSU	ck: 1929	len: 302	finds: 1	P54604 bacillus subtilis. hypc
SW:YHG3_YEAST	ck: 5569	len: 429	finds: 1	P38756 saccharomyces cerevisiae
SW:YHGN_ECOLI	ck: 1678	len: 197	finds: 1	P46891 escherichia coli. hypot
SW:YI29_MYCTU	ck: 7074	len: 164	finds: 1	Q50604 mycobacterium tuberculo
SW:YJ96_MYCTU	ck: 1836	len: 317	finds: 1	Q10862 mycobacterium tuberculo
SW:YJUX_ECOLI	ck: 3873	len: 173	finds: 1	P39411 escherichia coli. hypot
SW:YJUX_ENTAE	ck: 4167	len: 54	finds: 1	P39430 enterobacter aerogenes
SW:YJH8_YEAST	ck: 8234	len: 196	finds: 1	P46989 saccharomyces cerevisiae
SW:YKRF_ECOLI	ck: 8141	len: 79	finds: 1	P75677 escherichia coli. hypot
SW:YKZ7_MARPO	ck: 7679	len: 69	finds: 1	P38469 marchantia polymorpha
SW:YN06_YEAST	ck: 7226	len: 139	finds: 1	P53842 saccharomyces cerevisiae
SW:YNAL_CABEL	ck: 5825	len: 170	finds: 1	P34496 caenorhabditis elegans.
SW:YNC9_CABEL	ck: 399	len: 336	finds: 1	P34542 caenorhabditis elegans.
SW:YN06_CABEL	ck: 431	len: 381	finds: 1	P50444 caenorhabditis elegans.
SW:YORL_TTV1	ck: 9154	len: 232	finds: 1	P19296 thermoproteus tenax v1h
SW:YOTC_CABEL	ck: 9212	len: 161	finds: 1	Q10120 caenorhabditis elegans
SW:YOXD_BACSU	ck: 2171	len: 238	finds: 1	P14802 bacillus subtilis. hypc
SW:YFDC_ECOLI	ck: 486	len: 285	finds: 1	P77396 escherichia coli. hypot

```

\\End of list
Databases searched:
SWISS-PROT, Release 38.1, Released on 20Nov1999, Formatted on 28Dec1999

Total finds:      337
Total length:    29,864,866
Total sequences:   82,229
CPU time:        03:12.78

1 FINDPATTERNS on Swiss-Prots: * allowing 0 mismatches

1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C,T,S)(R,
AAKG_PIG CK: 8859 len: 133 i Q09138 sus scrofa (pig). 5'-amp-activate
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x(75)(A)x(P)(V)xx(Y)(S)(K)x(V)xxx(G)x(60)
LVFEDSLQVKRSALVOIYELEHKKXPVIDPSGNTILTLTKLTIEFFPKPEMSKSLDELGI
1:
ACF_BACSU CK: 8148 len: 77 i P80643 bacillus subtilis. acyl carrier P
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x(A)(T)(L,x)(V)(T)(K)xx(X)(A)xxx(T)x(92)
NADTLERVTKIYDRGLGVADADVAKLEAFERDLGDSDLVELVMELEDDEFDMETSDAE
1:
ADH2_DROMO CK: 8528 len: 253 i P09369 drosophila melanogaster (fruit fly)
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x(145)(P)x(P)(V)xx(A)(S)(K)xx(A)xxx(T)x(92)
AIANKNKITFAAGJGGSGFDSPSEIVSGSKNVILLIDRIENPAIAIELKALNPXYTVTFPPDY
1:
INGGITKTPLVHKFNLSMDLYEPRVAELLETHTOTLLQCQNFKAITANONGMIW
KLDGTLEALEMTKHMDSHI
ALR_TREPA CK: 255 len: 357 i Q56346 treponema pallidum. alanine racem
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x(184)(A)x(S)(A)x(L)(C)(H)xx(A)xxx(A)x(157)
MSRTARVCLEPVKDAYGHGCADVACALSCGSHRAVCGEASLRAGAAGRADILCSPTPT
1:
STPIYHANSAULICHPRAHDWRPGCLLAVGAPESVHAHVATLIPMELVLOVY
RAIKTIKGAIVSYORLMRAHTEHHVILDIGADVAMRALSPGOVLCGRWKYPVGAICMDCOVDLGTPLATVYGDNTLFG
AMIA_SALTU CK: 5663 len: 289 i P33772 salmonella typhimurium. probable
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W

```

x(22)(A)x(T)(L)xx(M)(S)(H)xx(A)xxx(T)x(251)
MSTFKLTLTSSRRQVLTGLAALTSGMSHAAVKEETLTSGNSHSPKTKGSKRLVMDPGHG
ATDRHLLQOVLFVDTGDTKNSLTGSLTKIKRSHRTTEQAAFLYKRS
PSIPSVLVEITSFIINPEERLGTAFAROKIAAIAANGISYFHWFDNOKAHTKKR

1:
AMP_C_SERMA ck: 107 len: 376 | P18539 serratia marcescens beta-lactamase
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(191)(M)x(P)(A)xx(M)(S)(H)xx(V)xxx(A)x(169)
MTKNRCAALIALILPTAHAAQODIDAVIOPLMKRYGVPGMAIAVSDGKOIIPYVASKQTC
KSLDOPFIOAMEOGLPALGNSHTYVPAQAQANVAQAGDEKPRVNPGLDA
ESYGISNARBLIRYLANILOOVAVASVARRPRTSVISAGAFODLWENYPPVPLSLRLEGNNAGMIMNGTAPATATPPQPEL

1:
AMPD_CITFR ck: 4831 len: 187 | Q00831 citrobacter freundii, and enterobact
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(164)(P)x(P)(A)xx(W)(S)(R)xx(A)xxx(T)x(7)
MLENGMLVDARHVPSPHDCRDEKFTLLVANHISLPPEFGGFWIDALFTGTIDPDHPFPAE
LTTSSDKET

1:
ANX2_XENLA ck: 6452 len: 339 | P24801 xenopus laevis (african clawed frog)
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(23)(L)x(S)(A)xx(A)(T)(H)xx(A)xxx(A)x(300)
ALIHETLKSLSEGNSSROSKSGVKAATHFPAEKDAAEIAETAKTGVDDELTIINILTRNSNE
BROEDGNMVDYERIDOBARELYEAGVARKGTMTTMTSESHPHLOKVERRYK
SYSYDIEESTIKREVKODLEAFNLTVOCIONNPLVFADLYESMGKGKXKILLIRVSRNMDMKIRPEFKKXKGLHFTIC

1:
ARG1_XENLA ck: 4321 len: 360 | Q91553 xenopus laevis (african clawed frog)
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(90)(P)x(T)(V)xx(A)(C)(R)xx(A)xxx(S)x(254)
MSIRNFEVRLKKOVSIILKOKKSHSAVAGAFSGQKRGVEHPAARSAGLIERLSNLGNC
PPRPGFSWAKFCLSKSDIYIGLDDPAPOFILKNNTISYSMRHIDCMGIRKVM
EKTFDQLGRDRPIHLSFDIDADPALATGTPVIGLTYREGVYITEEIHNTGMLSAVDLVEVNPVLAATSEEVKATANLAVDYI

1:
ARG2_XENLA ck: 4641 len: 360 | Q91554 xenopus laevis (african clawed frog)
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(90)(P)x(T)(V)xx(A)(C)(R)xx(A)xxx(S)x(254)
MSIRNFEVRLKKOVSIILKOKKSHSAVAGAFSGQKRGVEHPAARSAGLIERLSNLGNC
PPRPGFSWAKFCLSKSDIYIGLDDPAPOFILKNNTISYSMRHIDCMGIRKVM
EKTFDQLGRDRPIHLSFDIDADPALATGTPVIGLTYREGVYITEEIHNTGMLSAVDLVEVNPVLAATSEEVKATANLAVDYI

1:
ARG3_XENLA ck: 4724 len: 360 | Q91555 xenopus laevis (african clawed frog)
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(90)(P)x(T)(V)xx(A)(C)(R)xx(A)xxx(S)x(254)
MSIRNFEVRLKKOVSIILKOKKSHSAVAGAFSGQKRGVEHPAARSAGLIERLSNLGNC
PPRPGFSWAKFCLSKSDIYIGLDDPAPOFILKNNTISYSMRHIDCMGIRKVM
EKTFDQLGRDRPIHLSFDIDADPALATGTPVIGLTYREGVYITEEIHNTGMLSAVDLVEVNPVLAATSEEVKATANLAVDYI

1:
ARGC_ECOLI ck: 6342 len: 334 | P11446 escherichia coli, n-acetyl-gamma-gl
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(77)(L)x(T)(A)xx(V)(S)(R)xx(A)xxx(L)x(241)
MNLTLVAGASGAGAEIVTVNRRPHNNTITATLSASNSAGLISDLHQKGLYDLELQMSDI
EYVATSGVSGACLSKSDIYIGLDDPAPOFILKNNTISYSMRHIDCMGIRKVM
EKTFDQLGRDRPIHLSFDIDADPALATGTPVIGLTYREGVYITEEIHNTGMLSAVDLVEVNPVLAATSEEVKATANLAVDYI

1:
ATP8_PODAN ck: 8763 len: 50 | Q02653 podospira anserina, atp synthase
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(20)(A)x(T)(V)xx(L)(S)(K)x(L)xxx(V)x(14)
MPQVPEFVFEIIFFTIILAIYIILSKITLDRFRLSLRIFSKLGG

1:
ATP8_PONPA ck: 3336 len: 68 | P92694 pongo pygmaeus abelii (sumatran o
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(34)(P)x(T)(P)xx(F)(T)(K)xx(L)xxx(P)x(18)
MPQNTTWTPTIITPMLALFLITQLKLSHSHLPPPTPKFTKPKLHAKPWCMPKTKYLP

1:
ATP8_PONPY ck: 3550 len: 68 | Q35584 pongo pygmaeus (orangutan), atp s
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(34)(P)x(T)(P)xx(F)(T)(K)xx(L)xxx(P)x(18)
MPQNTTWTPTIITPMLALFLITQLKLSHSHLPPPTPKFTKPKLHAKPWCMPKTKYLP

1:
ATPD_HUMAN ck: 1686 len: 168 | P30049 homo sapiens (human), atp synthas
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(144)(V)x(T)(A)xx(A)(T)(K)xx(I)xxx(I)x(11)
MLPALLRRLRGIGLRLVHARAAVEAAAPAAASGPNMSTFASPTOVFENGANVROYDVPTL

1:
ATP2_BACP3 ck: 4392 len: 127 | P09354 bacillus p33 (thermophilic bacter
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(68)(V)x(T)(L)xx(L)(S)(R)xx(L)xxx(A)x(43)
MGNDAEMRQVRYITLILALITITIGFRTYKRVPLSTIGLSTISILMTWNTWKTEKGO

1:
BMPA_XENLA ck: 9714 len: 398 | P25703 xenopus laevis (african clawed fr
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(178)(V)x(P)(A)xx(A)(S)(R)xx(V)xxx(L)x(204)
MVAGISHLLELYQVLLSGCGGLIEBEGRKRYTSGSSPOOSQVRLNQPTELRLSFGIKR
IYVPAASRGPVRLDTRLVHNSKESSEVDYPAIARWIAIKROPHNGFVENV
HLNDKVPKRKHVIRSLTLPDKDMPQIRPLVTFSHDGKHALHKROKROARHKRRLKSSCRHRLPYVDFSVGNWMDIYA

1:
BMPB_XENLA ck: 1620 len: 398 | P30884 xenopus laevis (african clawed fr
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(178)(V)x(P)(A)xx(A)(S)(R)xx(V)xxx(L)x(204)
MVAGISHLLELYQVLLSGCGGLIEBEGRKRYTSGSSPOOSQVRLNQPTELRLSFGIKR
IYVPAASRGPVRLDTRLVHNSKESSEVDYPAIARWIAIKROPHNGFVENV
HLNDKVPKRKHVIRSLTLPDKDMPQIRPLVTFSHDGKHALHKROKROARHKRRLKSSCRHRLPYVDFSVGNWMDIYA

1:
BPHB_PSEPS ck: 1527 len: 275 | P08694 pseudomonas pseudocalleigenes, bl
<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(123)(A)x(P)(A)xx(A)(S)(R)xx(V)xxx(I)x(136)
MKLGEVAVLTGGASGLRGLADRFPAEKVAVLDSABRLALETDLGDNVLIGVDVSGLE
ATRGDAAPASGALVYNDGLVGRGFGSGAGNDLLEOLNIHP

1:
BR3A_HUMAN ck: 2610 len: 423 | Q01851 homo sapiens (human), brain-spect

1	CPCL_CANPG	ck: 1605	len: 114	1	P81580 cancer pagurus (rock crab). cuticle
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{54}(P)x(P)(P)xx(V)(S)(H)xx(A)xxx(S)x{44}) QVYSGIVSPDGNNOIOTHFHSHSYLKPSGIVTSDGNKQLTLAGQASLQAAAPAPLPVSHY			
1	CPTR_PEA	ck: 5889	len: 402	1	P21727 pisum sativum (garden pea). triose P
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{85}(P)x(P)(V)xx(F)(S)(R)xx(G)xxx(G)x{30}) MESVLSRATLISLPITLNLRLPLANASLPVSKFSGVSDDGNLVGRQLRPPLCSPYLKKGAS TSNVSPAFAVAFHTVALBEPFENMAASOPILCOISDITLMLSLAPVTVGSMA5 LTLSFNWIGFISAMISFTYRSYKRAMTMDSTNIYATISIALIYCIPPALIEGPPLTKTFPNDIAKVGIVKESVDLFWY			
1	CSCR_ECOLI	ck: 8969	len: 331	1	P40715 escherichia coli. sucrose operon rep
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24}(A)x(S)(V)xx(A)(T)(R)xx(V)xxx(I)x{29}) MASLKVAPALVSMATVSRVWNESTYRTRDQVDAIOTLNTVVDLSARKMRAGKRPSTLAY ROPLOCFWLPESALATGYRROGEADAMWRAGDLEAVQFPHATGDDHTULASLN AHFSGKDPFDVLICGNBRAAFVAYQVLLAGVRIPODVAVMGFDLVGVHLFPLPTTITQLPNDIIGREALHIISGREGVTRI			
1	CTAA_BACFI	ck: 716	len: 297	1	Q04443 bacillus firmus. cytochrome aa3 cont
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{52}(I)x(T)(L)xx(V)(S)(H)xx(V)GSGEGCATWPCFCEVLEPTNPATETIIEVSHRVSC MKRKIVYVITISIVLITLLOGLAVTSGSGEGCATWPCFCEVLEPTNPATETIIEVSHRVSC YKHSSEATLACAGFPLCGOIFPCIGVGVAGVHVPVGTILLITLIMITLSR YHRIKLVITAVLSLTVAGVOTISGISTVTONALSVGLHALITSLFSAISYMTMITRPSH			
1	CTAA_BACSV	ck: 1522	len: 306	1	P12946 bacillus subtilis. cytochrome aa3 co
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{51}(P)x(S)(L)xx(W)(S)(H)xx(A)xxx(S)x{23}) MNRKALCVLTFTFWLITLIGLAVTSGSGEGCATWPCFCEVLEPTNPATETIIEVSHRVASGI TGAVYHRESSLACGNVPLCGPLNNGPTORHEWONGHRAALLFVMTIIVAVH ATTSKDKQOIFMWSICLIFITLQALSGIMIVYSELATGFALHSHFICLGVLCYFILLARRRYSRS			
1	CU04_BLACR	ck: 1926	len: 127	1	P80675 blaberus craniifer. cuticle protein
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{103}(I)x(P)(L)xx(V)(S)(H)xx(V)xxx(V)x{8} QAVLYPSIYSILSKSVKVOLEPVEGAAPADTKKAEIKERHEVITHGPPVYSVVOPIVS			
1	CY2_RHOTE	ck: 2620	len: 92	1	P00098 rhodocyclus tenuis (Rhodospirillum rubrum) cytochrome c2
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{64}(P)x(P)(A)xx(V)(T)(K)xx(A)xxx(V)x{112}) ADESALAKOTKGLACNHPKAVVPAIKNVAKTIAGQAKAKLVAKYAGGQVMAKOLGAEI			
1	CYNR_ECOLI	ck: 4304	len: 299	1	P27111 escherichia coli. cyn operon transcr
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{192}(L)x(S)(A)xx(A)(T)(R)xx(I)xxx(C)x{91}) MLSHINTELAENHGSFTFRAASALHVSOPLOIRGLESLGVPLFDSGRTIRLTADGEVWRG HBOVALSRHDEKVLVLSAEFATREPTROIDHYCEKRGIDVETANSIVLELIRR TSLSTILPRAKITQHDGLKRAISLAPLERAVLLRRKNSQTAAKAFPLMALDKCAVVGNE5R			
1	DACX_STRSK	ck: 4817	len: 291	1	P39042 streptomyces sp. (strain k15). d-ala
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{23}(A)x(P)(A)xx(V)(T)(K)xx(I)xxx(G)x{252}) MRLRRAATVITTGALLAAGLIGATPATAVATKPTIAAAGVAMNNGTGLTYTKADTRSTGSTI			
1	ANSTPRHLTIASSAKNSTERTVYKTAIVTKTGSTIRMDTKNTGLLS				
1	SYSGAIGVKTGSGPEAKYCLVPAATRGKTVIGTIVASTISIPARESDATKIMNGFAL				
1	DBP5_YEAST	ck: 2958	len: 482	1	P20449 saccharomyces cerevisiae (baker's
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{146}(A)x(S)(L)xx(L)(T)(R)xx(P)xxx(S)x{1320}) MSDTRDPADLSTIKITDNEKTSVSTKQPEKPTADSIRPAKLVPRVKEKTRQED TLEVYQMGKFTITISQIIVPDSFEKRNINNOYVGPYVLDIMRKMLQLOKI KIFVLEADNNLDOGGDGCIAVKNFELDKDQVLSFTIRADAVKQIAKIVPNAITLQTNVAVDAIKQIYDCKNEADKF			
1	DEP_STRMO	ck: 9135	len: 145	1	Q54433 streptococcus mutans. dna/pantoth
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{45}(P)x(T)(L)xx(L)(S)(K)xx(V)xxx(V)x{84}) MTKIIILVSGSIAAYKAAADLSHQLTKLGYNHVMPTNAKQFIPLTLQVLSKNPVSYNWK			
1	DHAT_CITFR	ck: 101	len: 387	1	P45513 citrobacter freundii. 1,3-propane
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{143}(A)x(T)(A)xx(V)(T)(R)xx(V)xxx(T)x{228}) MSYMEFYLVPNVNFGRPNASIVSGRCKLGGKRALVTDGGLAIRIDGAVDRTLHLREAG INDEPLMLGRPAPLITATGMDALTHAVEVYISKDANPVTDAALQAIILIRANLRO AVALGSNIKARENNAYASLILAGMAFNANLGVHAMAHQGLXDMPHGVANAVLLPHVARYNLIANPEKFAIDIAFEGMENTDGL			
1	DHAT_KLEPN	ck: 846	len: 387	1	Q59477 klebsiella pneumoniae. 1,3-propan
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{143}(A)x(T)(A)xx(V)(T)(R)xx(V)xxx(T)x{228}) MSYMEFYLVPNVNFGRPNASIVSGRCKLGGKRALVTDGGLAIRIDGAVDRTLHLREAG INDEPLMLGRPAPLITATGMDALTHAVEVYISKDANPVTDAALQAIILIRANLRO AVALGSNIKARENNAYASLILAGMAFNANLGVHAMAHQGLXDMPHGVANAVLLPHVARYNLIANPEKFAIDIAFEGMENTDGL			
1	DHB3_HUMAN	ck: 3434	len: 310	1	P37058 homo sapiens (human). estradiol 1
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{193}(L)x(S)(M)xx(A)(S)(K)xx(V)xxx(S)x{101}) MGVLEDFPILGLVGLVCLVCKMRORSQHLFLFCALPSSSLRSMGMAVITGAGDIGRAYSFE NISSGALRPMPPLYSIYASAKAFVYCFSSKALNVEYRDGIIIOVLTIPVSTPMTK YLNKTKRTADEVYKESLKYVTIGABSCCLAHETIITILNRPRIEYSTAQRFLTRYSDYLRKNISNR			
1	DHB3_MOUSE	ck: 9317	len: 305	1	P70385 mus musculus (mouse). estradiol 1
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{189}(L)x(S)(L)xx(A)(S)(K)xx(V)xxx(S)x{100}) MEKLFIAAGFVGLVCLVCKMRORSQHLFLFCALPSSSLRSMGMAVITGAGDIGRAYSFE NISSGALRPMPPLYSIYASAKAFVYCFSSKALNVEYRDGIIIOVLTIPVSTPMTK YLNKTKRTADEVYKESLKYVTIGABSCCLAHETIITILNRPRIEYSTAQRFLTRYSDYLRKNISNR			
1	DHB3_RAT	ck: 3366	len: 306	1	O54939 rattus norvegicus (rat). estradiol
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{189}(L)x(S)(L)xx(A)(S)(K)xx(V)xxx(S)x{100}) MEKLFIAAGFVGLVCLVCKMRORSQHLFLFCALPSSSLRSMGMAVITGAGDIGRAYSFE NISSGALRPMPPLYSIYASAKAFVYCFSSKALNVEYRDGIIIOVLTIPVSTPMTK YLNKTKRTADEVYKESLKYVTIGABSCCLAHETIITILNRPRIEYSTAQRFLTRYSDYLRKNISNR			
1	DHBR_MOUSE	ck: 4788	len: 312	1	O70503 mus musculus (mouse). putative st
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{161}(V)x(S)(V)xx(V)(T)(R)xx(L)xxx(V)x{135}) MECAPAGGLVYVAGWASTAVLDFRFAFVOCHEYKSGKGFVOSVMPY ERSKGVILNITSASGMLPVPLLTIVSATAPVDFESOCHEYKSGKGFVOSVMPY LVATIKLAKIOPRLIDRPSAETFEVKSATVGFQITRTGTVIHSLMSGINSIMPRMWFKIIIMGFSKSLRNRYLKKRKN			


```

ESETSYKVCCLDGPVFKAOEVAL
  1:
  RAS1_DROME ck: 2533 len: 336 i Q27297 drosophila melanogaster (fruit fly)
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(45)(V)x(S)(A)xx(A)(T)(R)xx(L)xxx(P)x(275)
    MEKLTNVAQOESEEBEGPSTYKLGISITAKDILLOQASLHVESYANATKOLAIPIGLGCG
    NESEVLONVAFTRAHNSDOOTKLIOAGMLESRESYALIVDSAMALYRSDYIGSG
    ELAARONHIGLFLRMLORLADEFGVAVVITNOVTASLDGAPGMDAKPKIGHIMAHSSITRLYRKGETRICKIYDSPCLPESEA
    1:
  RBCR_CHRYI ck: 4295 len: 302 i P25544 chromatium vinosum, rubisco operon t
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(83)(L)x(S)(L)xx(V)(S)(R)xx(L)xxx(V)x(203)
    MHVSLRLRYEPAYAHNSYTBAAEELHISOPAYSMQVQOLDEGLSLFERLGKQVYITAGREV
    GENA1SLARLAETFPVMBREGSGTROMERFSEGCOTIRBGMONTREANRCONR
    SGIGLSVSLHITIELETRRLTLDVGEFPRQWILYTRKRLSLPAAGAFREYLSAARHCRLG
  1:
  RBS0_SOLTU ck: 4699 len: 181 i P10647 solanum tuberosum (potato). ribulose
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(V)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(160)
    MASSIVSAAVATRSNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPIN
    1:
  RBS1_LYCES ck: 5240 len: 181 i P08706 lycopersicon esculentum (tomato). r
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(V)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(160)
    MASSIVSAAVATRSNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPIN
    1:
  RBS1_PETSP ck: 2967 len: 180 i P04714 petunia sp. (petunia). ribulose bisp
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(139)
    MASSIVSAAVATRTNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPYGR
    1:
  RBS1_SOLTU ck: 4863 len: 181 i P26574 solanum tuberosum (potato). ribulose
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(160)
    MASSIVSAAVATRTNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPIN
    1:
  RBS2_LYCES ck: 4535 len: 180 i P07179 lycopersicon esculentum (tomato). r
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(159)
    MASSIVSAAVATRSNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPIN
    1:
  EGY
  RBS3_SOLTU ck: 4433 len: 181 i P33764 solanum tuberosum (potato). ribulose
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(5)(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(160)
    MASSIVSAAVATRSNVQAQSMVAPFTGLKSAASFPVTKNNNVDTISLANSNGRVRCMQVWPIN
    1:
  PEGY
  RBSK_SCHPO ck: 3214 len: 318 i O60116 schizosaccharomyces pombe (fission y
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
    x(169)(P)x(P)(A)xx(L)(S)(H)xx(I)xxx(A)x(133)
    .
    .
    .
  1:
  MINIVLGSMTDLYMRKICRSGGCTHGEPPDGFSTONGGKCANOVAVARLSNPADTVKSM
  HMISTYCATLYPNEHEAAILINQADSPRTLENVAIVASKILSTFGRKAVITITIGSO
  GAYITSANESALVSACKVAKVDITTAAGDITGAFNSIAHGOPLKDSLEFAKCSAIVYORKGAASIPSLLEVDSFNLRKT
  1:
  RBS_MUSAC ck: 2518 len: 180 i Q24045 musa acuminata (banana). ribulose
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(6)(V)x(S)(A)xx(F)(T)(R)xx(P)xxx(A)x(158)
    MYSSMWSAATFTBASPQSSMVAPFTGLKSAASAFVTRKPNDLHLPSNGRGVQCMKWP
    1:
  TGY
  REG5_DROME ck: 7903 len: 298 i Q94913 drosophila melanogaster (fruit fl
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(131)(A)x(S)(L)xx(L)(T)(R)xx(L)xxx(A)x(151)
    MTTAAVTLACCLLAFHIOISSSALITMFFLIRNEKMSHLSITFQOLYSVACKSTAAVGL
    DKAATGCGOSSELEPVAAALTSERSEKRLTPIVIRPDSPPVEEKMPIDRDE
    DPLTSSWSGRSAQAFQDRHNOULKAALRLHAPRPAIPDAEASVPATGVRNSDEPOA
  1:
  REHY_TORRU ck: 4850 len: 218 i P52574 tortula ruralis (moss). rehydin.
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(160)(L)x(S)(L)xx(A)(S)(R)xx(I)xxx(A)x(42)
    MGSQWALGDLVPIQADSDSTMGHKRYDCKNGMTIIFSHPDYPPVCTTELGIKIAYNPEFEK
    1:
  WQGEPPVVISPSVSDKAKOMFPOGWETVNLKALRMTFVD
  REP2_YEAST ck: 6596 len: 296 i P03872 saccharomyces cerevisiae (baker's
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(160)(L)x(S)(V)xx(V)(C)(R)xx(I)xxx(I)x(264)
    MODIEAKLTYKARTAYAVNDVOC(LR)XEMAPVDDIESKRSDLLRPGYIRMESLTT
    1:
    MODIEAKLTYKARTAYAVNDVOC(LR)XEMAPVDDIESKRSDLLRPGYIRMESLTT
    TVNNANNTNEPFSRTRSDTNAGAKIKLONTITTECPKAVPIKRRVAVRAGKRSNTRSV
  1:
  RFB_NERGO ck: 5086 len: 346 i P37761 neisseria gonorrhoeae, dtap-gluco
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(160)(P)x(S)(P)xx(A)(S)(R)xx(A)xxx(V)x(170)
    MOTEGRKNTLVGAGFPGSAVVRHIIQNTDRSVNLDKITAGNLESLTDIADNPRYAEQV
    1:
    ANQRTYRLEPSTVNSNNYGRQEPPEKLIPLMIALNSGKPLPYTGDAQIRMLF
    VEDHARALYOVTEGVETIYNGHNEKTMLEVTICALLELEPAEKPAVARREDLITFQDRPHGARVADAANKIRDLG
  1:
  RHL_MACMU ck: 4371 len: 416 i Q28849 macaca mulatta (rhesus macaque).
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(108)(L)x(S)(I)xx(A)(T)(R)xx(M)xxx(I)x(292)
    SSKYPRSVAGCGLMALILLEAALLLEFFELFYDASLEDOGLVASTOVQODLTVMAVGLGF
    1:
    FGLTYAMCLPRKLPGETEDKYOTITSPSTPLMLMFWPFLNALLNIYIER
    KNAVSTYALAVASVTAIVSLSLAHOGKINMTYMNALAGVALSACHVISHBPWIAVLGVALISIGAKCLPVCNRY
  1:
  RHO7_HUMAN ck: 2901 len: 227 i P52198 homo sapiens (human). rho-related
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(126)(L)x(T)(L)xx(L)(S)(R)xx(L)xxx(T)x(85)
    MEGOSRCRCITVAVGAEKGTALLOVFKADVPQSYVTFVYENTASFEIDKRRIELNMWDT
    1:
    TVASIGRGHRLRTRDSRGMQSRQSLSGRPDRNDEEIHKDRAKSCNLM
  1:
  RIB7_ARCFU ck: 4460 len: 219 i Q28272 archaeglobus fulgidus. putative
    <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
    x(101)(A)x(T)(L)xx(V)(S)(R)xx(P)xxx(V)x(102)
    MRPYVGVNVAASLDRKIDESDKQRLRISCEDLRIYDLRLRESAIVWGIGTVLADDPRLITVK
    1:
    PFIIGRDSPTVCDSESLKCRKIEKIERIEGEGAVVARRNR
  
```

RIPB_LUFICY CK: 429 len: 277 i Q00465 luffa cylindrica (smooth loofah) (sf)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(57)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(204)
 MKRFYLLAIFVAASIVAEADRELSSTSKFIDGLKALPNSGTYYNIITLLSSAGSASH
 1: AKRSRKIEGOTIERISKNOPSLATISKEWASKOIOLOAQTNGTFKTPPV
 ITDDGQREIITVTSKYTNKIOILLNRYKONVAFEDVYSAXH

RIPB_LUFICY CK: 7681 len: 250 i P22851 luffa cylindrica (smooth loofah) (sf)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(38)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(116)
 ANVSFISGSDSSYSKFTTALRKALPSKREKVSNIPLPLPSAGASHYILMOLSNDAKAITMAH
 1: NEVSPALSLSENAWSLSKOTLOAQTNGAFRFPVILIDKNGQREIITLASKY
 QIKDVNSKLLNKNONIA

RK21_SPTOL CK: 718 len: 256 i P24613 splinacia oleracea (splinach). 50s rib
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(39)(V)X(S)(P)XX(L)(S)(R)XX(L)xxx(A)x(201)
 MASRTLASGSSSCATKLTBONLNPILLNPSPKFSGVVSPSLSLPLPYAKRRRFOEIP
 1: TNAVAHVEBOLLDDREIVFVKTKKNNIRNIGHOPITRIKIGTGYEDYPAS
 TLEAVEVEKEEAEEAEAEANPV

RR8_MARPO CK: 6002 len: 132 i P06362 marchantia polymorpha (liverwort). d
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(73)(I)x(T)(L)xx(L)(S)(C)(K)xx(L)xxx(S)x(43)
 MNDTIANNITISIRNANLGIKIVQFPAVITITINIKILFQSGIDFIDNKNONTKIDILINKYO
 1: MNDTIANNITISIRNANLGIKIVQFPAVITITINIKILFQSGIDFIDNKNONTKIDILINKYO

RS13_SCHPO CK: 3858 len: 150 i P28189 schizosaccharomyces pombe (fission y
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(192)(I)x(S)(L)xx(L)(T)(K)xx(L)xxx(S)x(117)
 GRMHSKGIASSALPYVSPAMCKADSDVQIILFKFKMSQSGVGLRSHGIPQVREIT
 1: GRMHSKGIASSALPYVSPAMCKADSDVQIILFKFKMSQSGVGLRSHGIPQVREIT

RS3_ACHAX CK: 7310 len: 257 i P41117 acholeplasma axanthum. 30s ribosomal
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(192)(I)x(S)(L)xx(L)(T)(K)xx(L)xxx(S)x(117)
 MGRKVNIFGFRVAVIHDMSKMTADKTIYPALVAKEDAVIRKFKLNKYNNAASHVEIERLKELKYN
 1: MGRKVNIFGFRVAVIHDMSKMTADKTIYPALVAKEDAVIRKFKLNKYNNAASHVEIERLKELKYN

RS8_SYNY3 CK: 8926 len: 133 i P73307 synechocystis sp. (strain pcc 6803).
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(74)(I)x(T)(L)xx(V)(S)(K)xx(L)xxx(S)x(43)
 MASDTIDMLTRIRNACAVRSHSTQVPTTKMLSTAKVLSSEGTIEDYSTGEGIKMKLVLLKX
 1: MASDTIDMLTRIRNACAVRSHSTQVPTTKMLSTAKVLSSEGTIEDYSTGEGIKMKLVLLKX

RUVB_THERM CK: 3575 len: 324 i Q56214 thermus aquaticus (subsp. thermophil)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(146)(L)x(T)(A)xx(L)(S)(R)xx(L)xxx(S)x(159)
 MEDLAPRTIDEVIGORLKKOKRLVLEAKRKEPELLEHLPGLGKTTLAHYAHHLGAVNI
 1: MEDLAPRTIDEVIGORLKKOKRLVLEAKRKEPELLEHLPGLGKTTLAHYAHHLGAVNI

MDPALLGVRIITEEALEIGRSKRIARVAKRFRPRYRDPAOGEVETRRRALE
 AALALGIDELGKREDBRELLVLRIRAGVGCATLALSLSDPGLBEVHDPYLIRGGLKRTPRGRVATELVAHHLGIPVGPFI
 1: AALALGIDELGKREDBRELLVLRIRAGVGCATLALSLSDPGLBEVHDPYLIRGGLKRTPRGRVATELVAHHLGIPVGPFI

S27A_ECOLI CK: 54 len: 217 i P26428 escherichia coli. sigma cross-reacti
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(122)(V)x(T)(L)xx(L)(S)(R)xx(A)xxx(C)x(179)
 MKRIGVILSGCAVYDSETHENVLTLAISRSAGAACFAPDQOQDVINHLTGEMETRNVLIE
 1: MKRIGVILSGCAVYDSETHENVLTLAISRSAGAACFAPDQOQDVINHLTGEMETRNVLIE

IYVEDNKRIVTTPAYMLAQNIAEASGIDKLSRYLAE

SECY_RICPR CK: 3620 len: 433 i Q9ZCS5 rickettsia prowazekii. preproteain
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(119)(I)x(T)(S)(M)xx(L)(S)(R)xx(A)xxx(L)x(220)
 MGOEFSKRSKSLNITFLPLILICRGSFPIPIPGDSTALNSVAKENKOGFILLGNKLSG
 1: MGOEFSKRSKSLNITFLPLILICRGSFPIPIPGDSTALNSVAKENKOGFILLGNKLSG

SIXA_ECOLI CK: 9360 len: 161 i P76502 escherichia coli. phosphohistidin
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(199)(V)x(S)(V)xx(L)(S)(H)xx(L)xxx(L)x(146)
 MGVFIMRGGDALDASDSVPLFTTNGCDSRLMAMWLKQGVETIERVLSVPTLRACQILEEV
 1: MGVFIMRGGDALDASDSVPLFTTNGCDSRLMAMWLKQGVETIERVLSVPTLRACQILEEV

SIXA_HAEIN CK: 2974 len: 164 i P44164 haemophilus influenzae. phosphohi
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(106)(V)x(S)(V)xx(V)(S)(H)xx(L)xxx(L)x(142)
 MNIFIMRGEAEVMANSDKARHLTVISGKQAFLOGGCKLKHSTLVLINSIDRLIVSPVYRAQ
 1: MNIFIMRGEAEVMANSDKARHLTVISGKQAFLOGGCKLKHSTLVLINSIDRLIVSPVYRAQ

SPAL_SALTY CK: 7477 len: 432 i P39444 salmonella typhimurium. probable
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(120)(V)x(P)(P)xx(A)(S)(R)xx(V)xxx(L)x(296)
 MKTPILOVLAIVPQKITGPILIAELADVAIGELIIPWLFKTCGCTCAGGWLTAQTHRAA
 1: MKTPILOVLAIVPQKITGPILIAELADVAIGELIIPWLFKTCGCTCAGGWLTAQTHRAA

SPTR_SPIKE CK: 9527 len: 241 i P21625 spiroplasma melliferum. spiralin.
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(44)(I)x(T)(P)xx(V)(T)(K)xx(L)xxx(L)x(181)
 KRILSLIAVGSVAVGATSVACNKTSSNNLSRVTIAPATVAASTPKAVTPEIKTALLEN
 1: KRILSLIAVGSVAVGATSVACNKTSSNNLSRVTIAPATVAASTPKAVTPEIKTALLEN

SRP_CHLPS CK: 9648 len: 134 i P28164 chlamydia psittaci. sulfur-rich p
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(134)(I)x(S)(V)xx(W)(C)(R)xx(L)xxx(L)x(84)
 MSEGANSIGSVISLIQPGLEQVQVQSLINSVLGMCVRHIIINPIKTSKIVOSRAFOI
 1: MSEGANSIGSVISLIQPGLEQVQVQSLINSVLGMCVRHIIINPIKTSKIVOSRAFOI

SSNU_SALTY CK: 3263 len: 352 i P96069 salmonella typhimurium. secretion
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(132)(L)x(S)(V)xx(L)(C)(K)xx(L)xxx(M)x(204)
 MSERKTOPEKRLKQBRKRGQVYSIETSLGFOIALYLFHFTEKMLILIESITFTLOLV
 1: MSERKTOPEKRLKQBRKRGQVYSIETSLGFOIALYLFHFTEKMLILIESITFTLOLV

SSRL_HUMAN CK: 8652 len: 391 i P30872 homo sapiens (human). somatostati
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(118)(I)x(P)(L)xx(P)(S)(S)xx(A)xxx(G)x(187)
 MFPNGTASPSSSPSSPSSGSGGSGSGPAGAGADNBERGRNASONGTSLSGQSAIILISF
 1: MFPNGTASPSSSPSSPSSGSGGSGSGPAGAGADNBERGRNASONGTSLSGQSAIILISF

SSRL_MOUSE CK: 8110 len: 391 i P30873 mus musculus (mouse). somatostati
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(118)(I)x(P)(L)xx(P)(S)(S)xx(A)xxx(G)x(187)
 MFPNGTASPSSSPSSPSSGSGGSGSGPAGAGADNBERGRNASONGTSLSGQSAIILISF
 1: MFPNGTASPSSSPSSPSSGSGGSGSGPAGAGADNBERGRNASONGTSLSGQSAIILISF

NIGVAVLSLTVLPIVFSSTRANSNGTACNMIMPEAORMLVGFVLTITLMDGL

LPVGAICLCYVLIAMKRWALKAGQOKRBERITLTMVMMVVFVFCMPPYVQVLYNFAEDDADATVQSOLSVIILGANSKANP

SSRL_RAT ck: 9240 len: 391 i p28646 rattus norvegicus (rat). somatostati

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{186}(I)x(P)(I)xx(F)(S)(R)xx(A)xxx(G)x{187)
1: MFPNGTAPSTSPSSPBGCGVSRPGSGADMGEEGRNSQNGTLSEGGASIIISFIYS
NLGVMVLSLIVILPIVVFSTKANSNGTACNMILBEPNAGCMLVGFVLYTFELMGFL
LPGVAICLCYVLIAMKRWALKAGQOKRBERITLTMVMMVVFVFCMPPYVQVLYNFAEDDADATVQSOLSVIILGANSKANP

SSRB_CANFA ck: 8392 len: 183 i P23438 canis familiaris (dog). translocon-a

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{182}(I)x(P)(A)xx(V)(S)(H)xx(A)xxx(L)x{185)
1: MRLASVLAFLFVSHAEGBARILASKSLNNRYAVEGRDLTLOQINITNWSSAALDVELSDSFPF

PKSKRN

SSRB_HUMAN ck: 8093 len: 183 i P43308 homo sapiens (human). translocon-ass

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{182}(I)x(P)(A)xx(V)(S)(H)xx(V)xxx(L)x{185)
1: MRLSLFVLAFLFVTOAEGBARILASKSLNNRYAVEGRDLTLOQINITNWSSAALDVELSDSFPF

PKTKRN

ST14_SOLTU ck: 5285 len: 214 i Q41495 solanum tuberosum (potato). stsl4 p

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{43}(L)x(T)(A)xx(L)(T)(H)xx(A)xxx(P)x{155)
1: MEVSTMACTVLYIYIYIDEEKRRLKRYNNMTNLTFQGLLTITLTSLSLHSAQVPPPPPPPT

KKSIELEGCAQRTCEGPATLVCFNPNONVIGEKPY

SUFL_HAEIN ck: 4 len: 311 i P44847 haemophilus influenzae. sufi protei

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{122}(P)x(P)(L)xx(A)(S)(R)xx(L)xxx(P)x{273)
1: MPRLSRQLIKTAISTALSTVAPAPLALASREKLVPPLIEVRGRVITLTMQETNPILDGSHNVI
GVSDPIILIDOMEDNGLOLFQONDPHFVGRNLVNGIEIAPVLDVAGVTRILLL
NASLARAYDIRLNDDEMILODLFPKASVSVLSVSGEAEIIVNMKLTLYLSLAENVVACTRNKKYVLLR

SUW3_RAT ck: 800 len: 285 i P50235 rattus norvegicus (rat). alcohol su

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{111}(P)x(P)(A)xx(F)(S)(K)xx(L)xxx(C)x{1258)
1: MMSYTFEGEIPPAPEWSEKILLENCKFEVREKEDLIILTPRSNGNMLIETVCLORKGDPKWT
NEIUYVEDMKKTKSIRKICPGLPELIMLVKYSFOVYENNSNSL
MERKILITGLTPARKGTINDMKHFTVAQAEFADKVPDEKAGFPFGKFEPE

T2C1_CHV13 ck: 7233 len: 278 i P52283 chlorella virus ll-3a (cv-ll3a). cyF

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{146}(V)x(S)(M)xx(V)(S)(K)xx(P)xxx(T)x{114)
1: MEEKRKALIEKRIABEKIASGRKRIKSTINATKHEREREKYNIMFVGPTAFVVDIKNGKSR
AIFEGFDYCKPGRDNDVIGOGRPITTRKGRITLTLTFEGSALNGHLENFGEKHEP
VFYVTERSSSSGRSITTVNGVYKRLRFFIHPYNEVSSKTORI

TCF1_HUMAN ck: 5007 len: 269 i P36402 homo sapiens (human). t-cell-specific

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{91}(P)x(T)(V)xx(F)(T)(H)xx(L)xxx(S)x{1162)
1: MKTKTYSAPLNMHAPFVPSGAGQOPFOPPLHKNOPFHVPOLSLTYEHNSPHPPAPADISOR
LKSASAINOILGRMVALSREDDAKRERLHQLTPGWSARNDITGKKRR
SREHHESTETETWMPRELKDGNOESLSSSSSSPA

TELT_HUMAN ck: 3375 len: 167 i O15273 homo sapiens (human). telethonin (tl

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{139}(L)x(P)(V)xx(V)(S)(K)xx(A)xxx(S)x{12)
1: MATSELSEVSENEERDAEMAKDLITLSTRBEOCSLHEDTORHETHYHOGGOCQVLYOR

TELT_MOUSE ck: 5613 len: 167 i 070548 mus musculus (mouse). telethonin

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{139}(L)x(P)(V)xx(V)(S)(K)xx(P)xxx(T)x{12)
1: MATSELSEVSENEERDAEMAKDLITLSTRBEOCSLHEDTORHETHYHOGGOCQVLYOR

TRAM_AGR16 ck: 4597 len: 102 i O57471 agrobacterium tumefaciens. transc

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{113}(L)x(P)(L)xx(L)(T)(R)xx(P)xxx(L)x{73)
1: MELDANVTKKVLRPLIGITRGLPPTDLETITDAITHRILYERKADDELFOALPEYKKG

TRMD_SYNT3 ck: 6210 len: 231 i P72828 synechocystis sp. (strain pcc 680

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{70}(V)x(S)(L)xx(L)(S)(K)xx(V)xxx(T)x{145)
1: MODVITLPEPDFTSPLDLSGLGRKLEKALASVNLINPRDFTDKHRRVDEPDEGGVGMVTK
YTRPPEFRLAVPILLSGNHQAIAQWRLLEQOERTRQORRDLIMQKWRDPS

TRMD_SCHPO ck: 7997 len: 415 i O13947 schizosaccharomyces pombe (f1ssio

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{120}(I)xx(P)(L)xx(M)(T)(K)xx(L)(S)(K)xx(A)xx(A)x{199)
1: MRYSLFLOQITLKEKAFOPSTRLOAKORSQDKFYFVMSGVDSFSAVLKSGGVNVEGVEM
PTDKRQDITLCTIKLEAKLETITPRLNMTKEVVKROASAGCKEAEKESOGC
CFVSPNVRKRFELORYLNPDRIRIKYIAKNVYGECSNHCIMSLTVERGSLTLPQASEYRGMYWKIKNNALYICG

TRMU_YEAST ck: 8415 len: 417 i Q12093 saccharomyces cerevisiae (baker's

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{196}(L)(P)(I)xx(L)(T)(K)xx(V)xxx(A)x{205)
1: MLARYNLGRSAPFVORLPATFBDVIVYVAMSSGVDSVAALPAGFPNTRGYVOMNSE
PKDOSYLLSGINSTVLSLLPIGHLTPEDYRLADKAGLPTEKPEPSOGICFVNN
SOHGKRFNKLHYPSPGDIITVDPQSGAKTTGWRHDLGMSYTIQKVGISMPQADPNYQCTWFSKRLDTDEIILVGRDNP

TRNH_DATST ck: 5293 len: 268 i P50165 datura stramonium (jimsonweed) (c

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{162}(L)(S)(I)xx(A)(S)(I)xx(T)x{90)
1: MGRLEKEDRMSIRGCTALVGTGTRICAIYELANCALEYICSSQONDLDBCELEKWR
TKSLACEMAKDSRYVAVPWTINPTIEAACQVPSOKNIESLIGAPKRRAGEP
SEVSLVTLICLPTASTYITGLICVDGGTYNGEI

TRPF_CRYNE ck: 2342 len: 312 i P27170 cryptococcus neoformans (filobasi

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{62}(L)(P)(L)xx(V)(S)(K)xx(P)xx(I)x{234)
1: MSTSTVNALNGRDNVICALSGISHEDEKVEYKEGVALLGELMLBASDPKFLRSLLDL
GLDLYOHGDEPQAMKFPVAVVVFVPSGELVIRGEIIRGLNAILDLDGGA
SGGGEGKAPMEHAKRLTIOSEVGESEGVLPVILAGGLTPEVGOALNRLYKAGVMSAVGSEKREGSRRLRS

TRP1_CANFA ck: 4988 len: 371 i P43698 canis familiaris (dog). thyroid t

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{142}(L)x(S)(L)xx(V)(S)(K)xx(A)xxx(T)x{213)
1: MNSPKHRTFVSVDLSLSTLSEKRMKROAKDRADQOOLDOOS
RRFKQKYISAPEREHLASIMLITPTQVILFVONHRRKMKROAKDRADQOOLDOOS
GGGGGGGACGPOQOQOQSPRRVAVPVAVKPCQADAPAPGACVLDGHAQOQOQOQAAAAAISVSGSGGPGGLAHPG

1	TF1L_HUMAN	ck: 5439	len: 371	1	P43699	homo sapiens (human).	thyroid transd
							<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
							x{112}(L,I,V,M,A,P)X(P,T,S)(K)xx(A)xxx(S)x{213}
							MSMSPHHTPFYSVSLSESYKRGKMEGGGSLGAPLAAYRGGAPPTAAAOAHVGHGAVTA
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGCTGCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLVYKDGPCAGAPGAASLQGHAAQOQOAHQOAAQOAAAAISVSGAGLGAHPGHQF
							1: RRFKQOXYLSAPREHLASMIHLTPLOKINONRRKMRKOAADKAAOOLDOODS
							GGGGGGGGGGACCPQOQOAAQOOSRRRAVPLV

[illegible]

1	Y4TG_RHISN	ck: 6923	len: 231	1	P55661 rhizobium sp. (strain ngr234). probe
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{135}(L)X(P,A)X(F)(T)(H)XX(L)XX(1)X(80)
					MLYGTWDTGNGELFAISILPMLMGILITTLQAAFLGFFVACVLGWFVAVLRGMRTWVAMPAAV
					YEMLNAAKIIGDQTFNYLVLPLSMVGGIYLITLITVASALRIVDVNLPRKRGVPLR
1	Y503_METJA	ck: 6767	len: 406	1	O57926 methanococcus jannaschii. putative 2
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{160}(A)X(S)(A)XX(A)(T)(R)XX(1)XX(1)X(230)
					MTRVLMFMDLFEEFSMKRACVPYMPKLDKNDIYIDTTLRDGEOTPGVCTPKRQKLEIARKDELG
					VHKAEEAGADVRAHIDTGTGATPTQSMFEICKTKLEKKAHIGVCHDNFGEAVI
					NSIYGLIGAKAAVSTVNGIGERAGNMALEELIMALVLVDVLGNLEVLPELCRWVEYSGIKMKPKRPVGLVFAHESGIHVD
1	Y70A_METJA	ck: 4119	len: 102	1	P81311 methanococcus jannaschii. hypothetic
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{84}(1)X(S)(L)XX(V)(S)(H)XX(1)XX(L)X(2)
					MVGNNMRDKIKSIRKNWNIKRPITITVIGIVISAFAITISALMGMLFLITLITFSKITRKI
1	Y788_BORBU	ck: 1619	len: 440	1	O51728 borrelia burgdorferi (lyme disease s
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{164}(1)X(P)(L)XX(V)(S)(R)XX(1)XX(L)X(260)
					MHFLDENIOIKDKFYKNSLDKNVIAFEGADSTALNLKYYLSNVIAFYFAHFIRSDNEC
					NELSLNIGFVDSNTAONLTLARNRNNILPAIKFYKGYECLKRISEFKEFA
					DYFGDDEFFPEKCKYYSFDLKTFLDEPKYLVRLIFKILINSEGLAAKVSYKALNEAFKVEINRKNVNLKTNDFLEKRNKINI
1	Y79B_METJA	ck: 2865	len: 170	1	P81233 methanococcus jannaschii. hypothetic
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{19}(1)X(S)(L)XX(1)(T)(K)XX(L)XX(L)X(1135)
					MGINPFITIGMALILASIVSILILITLILFIDLAFGLSVGITTILITISIKILIKDKGRKKE
1	Y953_MCTU	ck: 883	len: 282	1	P71557 mycobacterium tuberculosis. hypothet
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{100}(1)X(T)(L)XX(L)(S)(H)XX(V)XX(V)X(166)
					MHYGLVLFSTDSGITPPAAALAAESHGRTFYVEHITHIPYKROAAHPTTGASLPPDDRYMRTLDF
					IPVLVGAAGTERKNFKIARASADGWITTPRDVIDDEPYKLAODIMAAAGRDGLPOIV
					ALDVVPYDCLKARNAELGVTEVLFGMPSRSDADNAAYVERLAKLACCV
1	Y98B_METJA	ck: 9981	len: 329	1	Q58395 methanococcus jannaschii. hypothetic
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{447}(A)X(S)(V)XX(L)(S)(R)XX(L)XX(G)X(1266)
					MELIETIKRDETLFCHNADPDAGVSCALKYLAQSOLNPKGRIRISADSVSLSNIIINEIGERY
					KDISOKLILYLSOESDVSKTALIKACSRRTEIEFKLIALISYSHSEASCACT
					IVSISADVAFYAAKKEKEIRVSARCKRKHVILGNLMEKIGKELGGSGGSHEDAGLNAFYDKSKSKKRVLKEVNLICRYFVE
1	YA32_PSEAE	ck: 6657	len: 122	1	P21485 pseudomonas aeruginosa. hypothetical
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{98}(P)X(S)(A)XX(A)(S)(R)XX(A)XX(1)X(8)
					MALASPACSRCCAAASARMPATACCPPTGASRTCARSRISPIWAMRWITCSTATVACASRRHMF
1	YABP_ECOLI	ck: 98	len: 216	1	P39220 escherichia coli. hypothetical 24.8
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{8}(P)X(T)(L)X(M,A)(S)(K)XX(1)XX(X)X(1992)
					MKVSVPGMPPVILLANSKRDYIKMVSQDKMDYKNNIFORIMETILRLHLEWSDKQDEAKYLENFVNNC
					SNDLIYERPGNANLGLPRLVRLPRTVLTITVLTWVLP
1	YECC_RHOCA	ck: 2127	len: 192	1	P26166 rhodobacter capsulatus (rhodospneu
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{142}(V)X(T)(L)XX(L)(T)(K)XX(W)XX(1)X(34)
					MDLLEFRAAHVAVEEMATHYIPEAARQIGAMWSDSRTIGFAOVITAIASRLQELHAIQTLVT
					VTNDLSLVISEFSLV
1	YBFT_BACSU	ck: 3368	len: 249	1	O31458 bacillus subtilis. hypothetical 2
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{184}(1)X(T)(T)XX(F)(S)(K)XX(V)XX(S)X(149)
					MKILIAHVEELCKLSAALIKEDIOAKRKADVLGATGSPVGLYKOLISDYQAGEIDSRYVT
					RLAISMGITIMEFSKHIVLLASGEKEKADAIOKMAEGPVTVDVPAISILQKRNHVTV
					IADYKAQRKLSASFS
1	YC08_KLEPN	ck: 4280	len: 373	1	Q48454 klebsiella pneumoniae. hypothetic
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{159}(A)X(S)(L)XX(L)(S)(K)XX(L)XX(V)X(1198)
					MNITLVNLTIPKIRIGAEVSVOLIESLIEKHSVTVSIEHNERKDTENGVKITYLPS
					YVGISNTIKDRIHAGEFFKSTERYTILYNSVKSIVLUDTLAANDKRLGFTGRLEK
					GFDOFCRLADLNKTRKFIAGEFDKNSASLQOLADSNVELGYCPVDPEKQVDIIVLPKMOEPGRVVEAIFPAKVYLTNR
1	YC21_METJA	ck: 2920	len: 299	1	Q58618 methanococcus jannaschii. hypothe
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{108}(1)X(T)(M)XX(1)(T)(K)XX(1)XX(1175)
					MMPEBRKRYMIIIPKFLNTVTEIKNNAYISIEPLKTSIEGLITITGNADARAERKIV
					LLCSGIYVCAFPSPSEFVKEVLEIYVLOIYSETSIISAIYAIAGSGMSISGKE
					YEIGWIVDSIILPILMGMALATDLYITFTILAINIVLVDVGYSIGLKVKYKINOKIKY
1	YC38_PORPU	ck: 4681	len: 291	1	P51321 porphyra purpurea. hypothetical 3
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{110}(L)X(T)(A)XX(1)(S)(R)XX(1)XX(L)X(1159)
					MTFAFSKRIKDPKPLKDPKXVLSIOTIEALVORFLQWRRPATIMACIOPLMVLYE
					VTMSLSTLTFGEHIELLALIVNLPFLPESSTALAPLPEPPEMLOLASLNLPIS
					YALGIRYIYSNTDMNFTESVIRKISWGDISTLGOIISLFLFDVIGAVYISNLKARLN
1	YCBQ_ECOLI	ck: 3224	len: 182	1	P75855 escherichia coli. hypothetical 61
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{444}(A)X(T)(L)XX(A)(T)(K)XX(V)XX(P)X(122)
					MITMKSVTLAFTIYVCATSSVMAADDNATDGSVFNKGVIAPACTIVAAATKDSVVLDPVS
1	YC6Z_ECOLI	ck: 1943	len: 78	1	P75991 escherichia coli. hypothetical 8.
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{17}(L)X(S)(A)XX(1)(T)(R)XX(A)XX(L)X(55)
					MHONSVIDLSGALITRIYFAKANLHIOETLGEIYVETILKDGRLNSRSLCAKLCLREHAT
1	YC01_ECOLI	ck: 6021	len: 262	1	P51983 escherichia coli. hypothetical 28
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x{116}(P)X(T)(1)X(V)(T)(R)XX(A)XX(1)X(230)
					MIIRKIDTTPVILPEBEETIMTYTPRRAEGCAPPTGTYHGSLLGAPLIMPAPAAHRESGL
					DPLACIEDPRLHSELGWLAQAFELPLVTVSGEYETPOSFGMCADNLICITAEFPF
					ISSDEASEKTLFAMANLRLRWHPKDAIRPS
1	YC1Z_ECOLI	ck: 51	len: 299	1	P77333 escherichia coli. hypothetical tr
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C

1: x[168](V)x(S)(P)xx(F)(S)(R)xx(V)xxx(V)x[115]
 MKREELIDMAFYVAERSTRAALMAQALSOIYRIEERLGLLRTTRSVPTAGEH
 RSVPTSVSOLIDHONINLYETSTANRMRLIGREVRERMEGOLLMTIDILID
 MAIDHGLATIPYDDEVERAKERKILIRLDEKFTPDGLYLYPHRRHNSAFSLFIDRLKTKYGV
 1: YCO5_YEAST len: 317 i P25616 saccharomyces cerevisiae (baker's ye
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[93](L)x(S)(V)xx(L)(T)(K)xx(L)xxx(S)x[100]
 x[15](L)x(T)(L)xx(I)(T)(K)xx(L)xxx(S)x[100]
 DPARAVKDMESSEOLELETKITKYLILKSLTYLEVENIEVIEHKAHLIEMIDYESTLV
 1: YD5A_SCHPO len: 174 i O14185 schizosaccharomyces pombe (fission y
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[97](L)x(S)(I)xx(F)(S)(R)xx(A)xxx(C)x[161]
 MTSQLEKAREWIEETLHRLKNOALDLDLOLGSYLIRICKKALGANIRKESNMFVOMENISA
 1: YDMG_SCHPO len: 344 i O14220 schizosaccharomyces pombe (fission y
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[46](L)x(S)(A)xx(L)(S)(K)xx(L)xxx(G)x[182]
 MNSIEINFTYLNLTISEVPEBNSVFTPKCSNSKLEHOSDEPSSACSLSKNTIIDGADH
 VLENHITSREPHETIKLPDYSLVNNMPHPPDYSKAPKDDFVPRFTGHS
 KLGCPICSHOGEFTWIKTSKATWIMNFVHGHKGRPPPIEFRTVLRKTRNAIGVNNKYMIEGCHQCKMKWIRCOGRDVS
 1: YER4_YEAST len: 128 i P40057 saccharomyces cerevisiae (baker's ye
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[49](L)x(P)(P)xx(F)(T)(K)xx(M)xxx(S)x[63]
 MEELICTPYHNSLMEFLFECSSKARAGHKPFLICYSNHLIPRLPPLSLFKRWALNPSS
 1: YF58_MYCTU len: 148 i Q10772 mycobacterium tuberculosis. hypotbet
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[7](L)x(S)(P)xx(W)(S)(R)xx(A)xxx(M)x[125]
 MFLSGEYAPSPDLMSREQADITMKSGTEGTOLOGFVILLITVGAKTGKTLKTKTILMVEHDGQVA
 1: YG29_BPSP1 len: 148 i P31653 bacteriophage sp01. hypothetical 16.
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[44](L)x(S)(V)xx(A)(T)(K)xx(M)xxx(P)x[89]
 MNOQWVLRIGVYVDPQPVKVESKYTGELVTRKVVHRLDIIIPVGLATKLTALSGALAAAG
 1: YGK9_YEAST len: 107 i P53138 saccharomyces cerevisiae (baker's ye
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[34](L)x(S)(L)xx(A)(T)(R)xx(L)xxx(S)x[57]
 MAQNPPLADIOVYKRYAKRMGOKKNSCTIAYIDSLOYCRSLSHKSCFPFPOSHAFSRPT
 1: YGIR_STROCO len: 66 i Q05954 streptomyces coelicolor. hypothetical
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x[12](L)x(S)(V)xx(A)(T)(R)xx(L)xxx(S)x[123]
 MAKVTRDVAKLKAGSTRVSVYVNNPBRVAVATERVLAIKELGVRPDRVAQMAASRTDL

1: YHAI_CRYPA len: 319 i P10941 cryptonecra parasitica (cheesnut
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[62](L)x(T)(P)x(V)(C)(K)xx(V)xxx(L)x[241]
 MAQLEKPSQSVISEYDPTTVPFVSRTVEEYVPGCITLWEYDSCGDPGLSHGDLRL
 1: YHCT_BACSU len: 302 i P54604 bacillus subtilis. hypothetical 3
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[24](L)x(T)(A)xx(A)(S)(K)xx(L)xxx(M)x[262]
 MNQKRGDELINERODQOMFVTLTKALASKPVLQDMWSHQOIKVNEESVNNMIVKQDR
 1: YHGN_ECOLI len: 197 i P46851 escherichia coli. hypothetical 21
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[120](L)x(T)(L)xx(L)(S)(H)xx(L)xxx(G)x[61]
 MNELISAVALIIMDPLGNLIFMSVILKHEPRRAIRVRELLIALVLYLFRAGEKILA
 1: YJ29_MYCTU len: 164 i Q50604 mycobacterium tuberculosis. hypot
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[49](L)x(P)(P)xx(L)(T)(H)xx(L)xxx(I)x[99]
 MGEYRVGIVNEQOPONPVYLLREANGDXYPLWIGOSSEMAALILEOGVBPPLTHDLIRD
 1: YJ96_MYCTU len: 1836 len: 317 i Q10862 mycobacterium tuberculosis. hypot
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[111](L)x(T)(L)xx(L)(S)(R)xx(A)xxx(L)x[190]
 MSAQGTINIGVVGDSFCSTHVAEMBAADQMRVAVLVGVVPTAPEGNAFEYSRFOE
 1: YJX3_ECOLI len: 173 i P39411 escherichia coli. hypothetical 18
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[144](L)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x[113]
 MIMHOVCATNPAKIOATILQAHHEIFGSGCHIASVAESGVPPQPGSEETAGARNVA
 1: YJX4_ENTAE len: 54 i P39430 enterobacter aerogenes (aerobacte
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[24](L)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x[113]
 EALGPVMSQHTGIDQIGREGAIGVFTACKLTRSSVHOAVVALLSPHNAIIR
 1: YJX8_YEAST len: 196 i P46989 saccharomyces cerevisiae (baker's
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
 x[167](L)x(S)(L)xx(F)(C)(K)xx(V)xxx(L)x[13]
 MLCGLDVLIPGDAITIIDDFDNISGFNVEETESALITLLKGAWTGANSDEKLEFOCNDN
 1: YVRIIRGSTAORGISAV len: 196 i P46989 saccharomyces cerevisiae (baker's

[illegible]

Mon Feb 14 08:07:43 2000

sp.cat

Page 25

RIKLOPVPLOSSENIKEAPSEKKEIKETNSFVLMNMTIMALMGVSYSEFP
IYSRFLGDEIGENILITRICIGRGIGALVSKNGFNNNGLFTFTVLISIALFLFPIPAVSIAALLFFIAMEYGEVLAKVKKQ
1 YVVP CAEEL ck: 1176 len: 274 : Q93834 caenorhabditis elegans. hypothetical
<(x){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(14){1}(T)(T)xx(F)(C)(K)xx(V)xxx(V)x(244)
1: MOGNAVEOEKIDQVYVYIMTPEFCCKWPTLAVGVDHDLTKSTNKIDGMELSLRPTPELPKRVVA
AGAVVYVMTQTSIESMTPLATSPISMASKMKFGKRGGLVYRPERVSTWHR
SYAIKEGFEOKVMPDVYNAFYHRIKMKPEPTEFRSEELL
ZRP4_MAIZE ck: 6645 len: 364 : P47917 zea mays (maize). o-methyltransferase
<(x){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(99){V}x(T)(L)xx(V)(S)(R)(L)xx(S)x(249)
1: MELSPNNSTQSLDQLEHMTPEKSKMALKRAIHRADAIHLHGAAASLSQILSKVHLHP
GLASDQSLIVDAIKQSAEYFGISSLVGVGGIGAAVAISKAPFKVCYLDL
HVAKAPLTHDVQFIAGDMFESIPADAVALLKSVLHMDHDCVILKNCKALPFRAGKVIITIMVYGAGPSDMKHEKMOALFDV
Databases searched:
SWISS-PROT, Release 38.1, Released on 20Nov1999, Formatted on 28Dec1999
Total finds: 337
Total length: 29,864,866
Total sequences: 82,229
CPU time: 03:13.70
1:AA_SEQUENCE 1.0 STANDARD: PRT: 133 AA.
ID ACP_PIG
AC Q09138 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE 5'-AMP-ACTIVATED PROTEIN KINASE, GAMMA-1 SUBUNIT (AMP GAMMA-1 CHAIN)
DE (38 KD SUBUNIT) (FRAGMENTS).
GN PRKAG1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RC TISSUE=LIVER;
RP SEQUENCE.
RX MEDLINE: 95050763.
RA STABLETON D., GAO G., MICHELL B.J., WIDMER J., MITCHELL K.,
RA TEH T., HOUSE C.M., WITTERS L.A., KEMP B.E.;
RT "Mammalian 5'-AMP-activated protein kinase non-catalytic subunits are
RT homologs of proteins that interact with yeast Sn1 protein kinase.";
RT J. Biol. Chem. 269:29343-29346(1994).
CC -1- FUNCTION: AMPK IS RESPONSIBLE FOR THE REGULATION OF FATTY ACID
CC SYNTHESIS BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO
CC REGULATES CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND
CC INACTIVATION OF HORMONE-SENSITIVE LIPASE AND
CC HYDROXYMETHYLGUANYL-COA REDUCTASE. APPEARS TO ACT AS A METABOLIC
CC STRESS-SENSING PROTEIN KINASE SWITCHING OFF BIOSYNTHETIC PATHWAYS
CC WHEN CELLULAR ATP LEVELS ARE DEPLETED AND WHEN 5'-AMP RISES IN
CC RESPONSE TO FUEL LIMITATION AND/OR HYPOXIA. THIS IS A REGULATORY
CC SUBUNIT.
CC -1- SUBUNIT. HETEROTRIMER OF AN ALPHA CATALYTIC SUBUNIT, A BETA AND A
CC GAMMA NON-CATALYTIC REGULATOR SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE 5'-AMP-ACTIVATED PROTEIN KINASE, GAMMA
CC SUBUNIT FAMILY.
CC -1- SIMILARITY: CONTAINS 4 CBS DOMAINS.
DR PFAM: PF00571; CBS 1.
KW Fatty acid biosynthesis; Repeat; CBS domain.
FT NON_TER 1
FT DOMAIN <1 >11 CBS 1.
FT DOMAIN <26 >42 CBS 2.
FT DOMAIN 62 91 CBS 3.
FT DOMAIN <95 127 CBS 4.
FT NON_CONS 11 12

FT NON_CONS 24 25
FT NON_CONS 42 43
FT NON_CONS 73 74
FT NON_CONS 80 81
FT NON_CONS 94 95
FT NON_CONS 103 104
FT NON_TER 133 133
SO SEQUENCE 133 AA: 14763 MW: 5D039F44 CRC32:
AANG_PIG Length: 133 February 14, 2000 08:02 Type: P Check: 8859
1 LVEDTSLQV KSAIVQIEL EHKXPVIDP ESGNTIYILF XLFITEPCK
51 PEFMSKSLSE LQIGTYANIA MVRVSLPVY IYSEFVINYL AAEKSHYFEG
101 YKLVAVDEN DVKGIYSLS DILQALVYLNG GEK
1:AA_SEQUENCE 1.0 STANDARD: PRT: 77 AA.
ID ACP_BACSU
AC P80643; P51832.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE ACYL CARRIER PROTEIN (ACP).
GN ACPA OR ACP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE OF 15-77, AND SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=168;
RX MEDLINE: 96326321.
RA MORBIDONT H.R., DE MENDOZA D., CRONAN J.E. JR.;
RT "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RT lipid biosynthesis genes.";
RT J. Bacteriol. 178:4794-4800(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA OGURO A., KAKESHITA H., TAKAMATSU H., NAKAMURA K., YAMANE K.;
RT Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RX MEDLINE: 94131947.
RA HEATON M.P., NEUBAUS F.C.;
RT "Role of the D-alanyl carrier protein in the biosynthesis of D-alanyl-
RT lipoteichoic acid.";
RT J. Bacteriol. 176:681-690(1994).
CC -1- FUNCTION: THIS PROTEIN IS THE CARRIER OF THE GROWING FATTY ACID
CC CHAIN IN FATTY ACID BIOSYNTHESIS.
CC -1- PATHWAY: KEY COMPONENT IN DE NOVO FATTY ACID BIOSYNTHESIS.
CC -1- PTM: THE GROWING FATTY ACID CHAINS ARE COVALENTLY BOUND TO THE
CC 4'-PHOSPHOPANTETHEINE PROSTHETIC GROUP.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: U59433; AAC44308.1;
DR EMBL: D64116; BAA10975.1;
DR EMBL: Z98112; CAB13465.1;
DR HSRP: P03901; IACP.
DR SUBTILIST; BG11536; ACPA.
DR PROSITE; P580012; PHOSPHOPANTETHEINE. 1.
DR PROSITE; P580075; ACP DOMAIN; 1.
DR PFAM; PF00550; PP-binding; 1.
KW Fatty acid biosynthesis; Phosphopantetheine.
FT BINDING 33 AA: 8591 MW: 48D5166 CRC32:
SO SEQUENCE 77 AA: 8591 MW: 48D5166 CRC32:

ACP_BACSU Length: 77 February 14, 2000 08:02 Type: P Check: 8148 ..

1 MADLERVTK IYDRIGVDE ADVKLEASF EDGASLDV VELVLELDE
51 FDMESDEDA EKATVGDV NYIONQ

11AA_SEQUENCE 1.0 STANDARD; PRT: 253 AA.
ID ADH2_DROMO
AC P09369;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALCOHOL DEHYDROGENASE 2 (EC 1.1.1.1).
GN ADH2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RA MEDLINE: 89337903.
RA ATKINSON P.W., MILLS L.E., STARMER W.T., SULLIVAN D.T.;
RT "Structure and evolution of the Adh genes of Drosophila melanogaster";
RL Genetics 120:713-723(1988).
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37276; AAA28334.1; -
DR EMBL: X12536; CAA31054.1; -
DR PIR: S01901; S01901
DR EMBASE: F89n0012567; Dmc0\Adh2.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
DR PFAM: PF00663; adh_short; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0
FT NP_BIND 9 32 NAD (BY SIMILARITY).
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 253 AA; 27382 MW; 0136EC76 CRC32;
ADH2_DROMO Length: 253 February 14, 2000 08:02 Type: P Check: 8528 ..
1 AIAANKNIIV AGIGIGFDT SREIYKSGPK NLVILDRIN PAIAELKAL
51 NPKVTVTFYP YDVTVSAET TKLKTIPDK LKTVDLLING TGILDDHOIE
101 RTIAVNTGT LNTTAIMSF WDKRKGPGG VIANICVYG FNAIIPVYV
151 SASRAALSF TNSIARLAPI TGVAYSINP GITKTLVHK FNSMIDVEPR
201 VAEILLEHPT OTTIOCAQNF VKAIOANNG AIWKIDGLT EAIETKHMND
251 SHI
11AA_SEQUENCE 1.0 STANDARD; PRT: 357 AA.
ID ALR_TREPA
AC Q56346;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ALANINE RACEMASE (EC 5.1.1.1).

GN ALR OR TP0681.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN [1]
RP SPOUNCE FROM N.A.
RC STRAIN-NICHOLS;
RA MEDLINE: 98332770.
RA FRASER C.M., NOKRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM R.A.,
RA SODERREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388(1998).
RN [2]
RP SEQUENCE OF 1-178 FROM N.A.
RC STRAIN-NICHOLS;
RA STEINER B.M., RODES B.;
RT Partial sequence of alanine racemase from Treponema pallidum";
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-ALANINE -> D-ALANINE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001242; AAC65644.1; -
DR EMBL: U57756; AAB17466.1; -
DR TIGR: TP0681; -
DR PROSITE: PS00395; ALANINE_RACEMASE; FALSE_NEG.
DR PFAM: PF00842; Ala_racemase; 1.
KW Isomerase; Pyridoxal phosphate.
FT BINDING 13 13 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT SEQUENCE 357 AA; 38098 MW; 88BD7754 CRC32;
ALR_TREPA Length: 357 February 14, 2000 08:02 Type: P Check: 255 ..
1 MSRTARVCL PYKADAYGHG ACQVAQALS CGVHSFAVAC VQEAQOLRAA
51 GVRAPILCLST TPTAEISLS IEHRVHTVIS ERAHIALIAR ALROSDATGA
101 TCGVHWKIDP GKGRIQCAPD EACALVQWVC APTGLHLCGV CTRHSVADSV
151 RAEDLQYTEM QRAHMHQVQ YIKSGISIP LVHANSAL LCHRAHFDN
201 VRPGLATGY APESVPAVR SVLPVMEVY TVYRAIKIIP AGAVYSIQRL
251 WRAHTEHWG IIPIGYADV MRALSPGLOV CIGKWPYVY GAICMOCVY
301 DLGTPLRVTV GDRVTLFSPQ DAGGPGOGAD VLAHAGTIP YELCAIGKR
351 VERYIIR
11AA_SEQUENCE 1.0 STANDARD; PRT: 289 AA.
ID AMIA_SALTY
AC P33772;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE N-ACETYLURAMOYL-L-ALANINE AMIDASE AMIA PRECURSOR
DE (EC 3.5.1.28).
GN AMIA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2:
RX MEDLINE: 93352403.
RA XU K., ELLIOTT T.;
RT "An oxygen-dependent coproporphyrinogen oxidase encoded by the hamf
RU gene of Salmonella typhimurium."
U Bacteriol 175:4990-4999(1993).
CC -1- CATALYTIC ACTIVITY: HYDROLYZES THE LINK BETWEEN N-ACETYLURAMOYL
CC RESIDUES AND L-AMINO ACID RESIDUES IN CERTAIN BACTERIAL CELL-WALL
CC GLYCOPOLYMERES.
CC -1- SIMILARITY: STRONG, TO B. SUBTILIS LYSC/CMLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L19503; AAA27138.1;
DR STRYGENE; SG10008; AMIA.
DR PFAM: PF01520; Amidase_3; 1.
KW Hydrolyase; Cell wall; Signal.
FT SIGNAL 1
FT CHAIN 35 289
FT SEQUENCE 289 AA; 31659 MW; 5F107FEC CRC32;
SO AMIA_SALTY Length: 289 February 14, 2000 08:02 Type: P Check: 5663 ..

1 MSTRFLKLT TSRQVLYKTG IALTLGSGS HAVAKEETLK TENGSKPKT
51 KTKSGKRLVM LDPHGCGIDT GAIGNGSGE KHVYLATAKN VAIIRNNGI
101 DARLTRIGDT FIPLYDRVEI AKHGADLEF SIHGDGTNP KAAGASVFL
151 SNRGAASSMA KYLSRENRA DEVAGKRAKD RDHLQGVLF DLYQDTLKN
201 SITLGSHTLK KIKPIHKLHS RTTEQAAVYV LKSPISYVL VETSPITNE
251 EERLLGTATF ROKTATAIAN GIISIFHWF NOKAHTKKR

11AA_SEQUENCE 1.0
ID AMPC_SERMA STANDARD: PRT: 376 AA.
AC P18539;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-LACTAMASE PRECURSOR (EC 3.5.2.6) (CEPHALOSPORINASE).
GN AMPC.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OC [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SR50:
RX MEDLINE: 91032898.
RA NOMURA K., YOSHIDA T.;
RT "Nucleotide sequence of the Serratia marcescens SR50 chromosomal ampC
RU beta-lactamase gene."
U FEMS Microbiol Lett 58:295-299(1990).
CC -1- FUNCTION: THIS PROTEIN IS A SERINE BETA-LACTAMASE WITH A SUBSTRATE
CC SPECIFICITY FOR CEPHALOSPORINS.
CC -1- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O -> A SUBSTITUTED BETA-
CC AMINO ACID.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLASS C BETA-LACTAMASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52964; CA37137.1;
DR PIR: S11710; QKSE.
DR PIR: A48176; A48176.
DR HSSP: P05364; IBL5.
DR PROSITE: PS00336; BETA-LACTAMASE_C_1.
DR PFAM: PF00144; beta-lactamase; 1.
KW Hydrolyase; Antibiotic resistance; Periplasmic; Signal.
FT SIGNAL 21
FT CHAIN 22 376
FT ACT_SITE 79 79
FT BINDING 328 330
FT SEQUENCE 376 AA; 41096 MW; 016BEF00 CRC32;
SO AMPC_SERMA Length: 376 February 14, 2000 08:02 Type: P Check: 107 ..

1 MTKMNRCAAL IAAIIPTAH AAQOODIDAV IOPLMKRYGV PGMAIVASVD
51 GKQQLPYPGV ASKQTKPIT EQLTEVGSIL SKTFATLAV YAOQOSKLSF
101 KDPASHYLPD VRGSAFDGVS LNLATHTSG LPLFVPDDVY NNAQMLAYYR
151 ANQPKHPAGS YRYVSNLIGS MEGMTAAKSL DQPIQAMEQ GMLPALGMSH
201 TYQVPAAM ANVAGYSKD DKPVAVNPGP LDASVGIKS NNRDLIRYLD
251 ANIQVKYAS VARRWPRTS VITSAGFTQ DLMENVPYP VLSRLIEGN
301 NAGMNGTTP ATAITPPOPE LRAGWYKNTG STGFSFYAV FIPAKNIAYE
351 MANKWFPND DVEPAVHII QALEKR

11AA_SEQUENCE 1.0
ID AMPD_CITFR STANDARD: PRT: 187 AA.
AC Q00831;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AMPD PROTEIN.
GN AMPD.
OS Citrobacter freundii, and Enterobacter cloacae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OC [1]
RP SEQUENCE FROM N.A.
RX STRAIN-OS60, AND 14:
RX MEDLINE: 9319292.
RA KOPP U., WIEDMANN B., LINDQUIST S., NORMARK S.;
RT "Sequences of wild-type and mutant ampD genes of Citrobacter freundii
RU and Enterobacter cloacae."
U Antimicrob Agents Chemother 37:224-228(1993).
CC -1- FUNCTION: PUTATIVE SIGNALING PROTEIN IN BETA-LACTAMASE
CC REGULATION. AMPD SEEMS NOT TO ACT AS A DIRECT SENSOR FOR
CC BETA-LACTAMS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z14002; CA478390.1;
DR PIR: A48901; A48901.
DR PIR: S26138; S26138.

```

DR PIR: S26139; S26139.
 SO SEQUENCE 187 AA; 20839 MW; 229DF9F2 CRC32;
 AMPD_CITER Length: 187 February 14, 2000 08:02 Type: P Check: 4831 ..

1 MLENGWLIVD ARHVPSPHHD CRPEDEKPTL LVVHNISLPP GEGGSPWIDA
 51 LFTGIIPDA HPFPAFIHL ALSADCLIR DGEVQYVPE DKRAMHAGVS
 101 MYGRCRND FSIGLEGT DTPYTDAY EKLVAQTQL IGRPAIDN
 151 ITGSHDAPE RKTDPAPAD WSRFHMLTT SSDKEIT

11AA_SEQUENCE 1.0 STANDARD; PRT: 339 AA.
 ID ANX2_XENLA
 AC P24801;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE ANNEXIN II TYPE II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN)
 DE (CHROMOBINDIN 8) (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN
 DE IV) (PAP-IV).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-OOCYTE;
 RX MEDLINE: 92011609.
 RA IZANT J.G., BRYSON L.J.;
 RT "Xenopus annexin II (calpactin I) heavy chain has a distinct amino
 RT terminus".
 RL J Biol. Chem. 266:18560-18566(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 92009222.
 RA GERKE V., KOCH W., THIEL C.;
 RT "Primary structure and expression of the Xenopus laevis gene encoding
 RT annexin II".
 RL Gene 104:259-264(1991).
 CC -1- FUNCTION: CALCIUM-REGULATED MEMBRANE-BINDING PROTEIN WHOSE
 CC AFFINITY FOR CALCIUM IS GREATLY ENHANCED BY ANIONIC PHOSPHOLIPIDS.
 CC -1- BINDS TWO CALCIUM IONS WITH HIGH AFFINITY.
 CC -1- SUBUNIT: Tetramer of 2 light chains (P10 proteins) and 2 heavy
 CC chains (P36 proteins).
 CC -1- SUBCELLULAR LOCATION: IN THE LAMINA BENEATH THE PLASMA MEMBRANE.
 CC -1- TISSUE SPECIFICITY: ADULT BRAIN, HEART, STRIATED MUSCLE, LIVER,
 CC KIDNEY, AND VERY HIGH LEVELS IN SKIN.
 CC -1- DEVELOPMENTAL STAGE: THROUGHOUT OOGENESIS AND IN MATURE EGGS.
 CC CONSTANT LEVELS DURING EARLY EMBRYOGENESIS, BUT DECREASE AT 8H
 CC AFTER MIDLASTULA TRANSITION, THE STEADY STATE LEVEL INCREASES
 CC SUBSTANTIALLY.
 CC -1- DOMAIN: CONTAINS FOUR HOMOLOGOUS REPEATS WITH A CONSENSUS
 CC SEQUENCE COMMON TO ALL ANNEXIN PROTEINS. A PAIR OF THESE REPEATS
 CC MAY FORM ONE BINDING SITE FOR CALCIUM AND PHOSPHOLIPID.
 CC -1- MISCELLANEOUS: IT MAY CROSS-LINK PLASMA MEMBRANE PHOSPHOLIPIDS
 CC WITH ACTIN AND THE CYTOSKELETON AND BE INVOLVED WITH EXOCYTOSIS.
 CC -1- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: M58575; AAA49664.1;
 DR EMBL: M58575; AAA49665.1;
 DR EMBL: M60769; AAA49886.1;
 DR PIR: J01298; J01298.

DR PIR: A41002; A41002.
 DR PIR: B41002; B41002.
 DR HSSE; P04083; 1809.
 DR PROSITE; PS00223; ANNEXIN; 4.
 DR PFAM; PF00191; annexin; 4.
 KW Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation.
 FT INIT MET 0 0
 FT DOMAIN 1 24
 FT REPEAT 42 102
 FT REPEAT 114 174
 FT REPEAT 199 259
 FT REPEAT 274 334
 FT MOD_RES 26 26
 FT VARIANT 243 243
 FT VARIANT 257 257
 FT CONFLICT 222 223
 FT CONFLICT 297 297
 FT SEQUENCE 339 AA; 38643 MW; 01CF92AC CRC32;
 SO SEQUENCE

ANX2_XENLA Length: 339 February 14, 2000 08:02 Type: P Check: 6452 ..

1 ALIHEILKRL SLEGNQSSR QSKIGSVKAA THPDAEKDAA ALETAKTKG
 51 VDELTITNLT TNRSNEGRD IAFAPRRRK KDLPSALKGA LSGNLETWML
 101 GLIKTRPOYD ASELKAMKG LGTDEDLIE IICSTNREL LDIONAVREL
 151 FTELEKDIM SDTSGDPRKL MYALAKGRQ EDGNMVDYER IDQARELYE
 201 AGVRRKGTDV TKWTIMTER SHPHLOKVEE RYKSPSYDI ESIKKEVKG
 251 DEENAFMLNV QCIQNKPLYE ADRLYSKMG KGTMDKILIR IWSRNLDMD
 301 LKIRQEFKKK YKSLHFFIG QDTKGDYORA LNLNCGSDD

11AA_SEQUENCE 1.0 STANDARD; PRT: 360 AA.
 ID ARG1_XENLA
 AC Q91531;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ARGINASE, NON-HEPATIC 1 (EC 3.5.3.1).
 GN ARG1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 95014323.
 RA PATTERSON D., SHI Y.-B.;
 RT "Thyroid hormone-dependent differential regulation of multiple
 RT arginase genes during amphibian metamorphosis".
 RL J. Biol. Chem. 269:25328-25334(1994).
 CC -1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED
 CC IN TISSUE REMODELING.
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.
 CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN TAIL,
 CC INTESTINE, HINDLIMB AND TRUNK REGION. MOST ABUNDANT IN TADPOLE
 CC TAIL.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN NEURULA (STAGE 16/17). THE
 CC HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48. IN THE
 CC INTESTINE, INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS (STAGES
 CC 58-64). LOW LEVELS EXPRESSED IN HINDLIMB UNTIL STAGE 66 AFTER
 CC WHICH, LEVELS DRAMATICALLY INCREASE. IN THE TAIL, A CONSTANT HIGH
 CC LEVEL OF EXPRESSION IS FOUND THROUGHOUT METAMORPHOSIS.
 CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).
 CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
 CC -----
 CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: 008406; AA056891.1; -

DR HSBP; P07824; 2RLA.

DR PROSITE; PS00147; ARGINASE_1; 1.

DR PROSITE; PS00148; ARGINASE_2; 1.

DR PROSITE; PS01053; ARGINASE_3; 1.

DR PFM; PF00491; arginase; 1.

KW Urea cycle; Arginine metabolism; Hydrolase; Manganese; Multigene family.

FT METAL 122 122 MANGANESE 1 (BY SIMILARITY).

FT METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).

FT METAL 147 147 MANGANESE 2 (BY SIMILARITY).

FT METAL 149 149 MANGANESE 1 (BY SIMILARITY).

FT METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).

FT METAL 255 255 MANGANESE 2 (BY SIMILARITY).

SEQUENCE 360 AA; 39155 MW; 31E02C27 CRC32;

ARG1_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4321 ..

1 MSRSNPFRL LKQVSIIRL OKKSHSVAV IGAFSGKGR RGVHGPAA

51 IRSAGLIERL SNLCNVCDF GDHFSGVFN DELYSIVKH PRVGLACKV

101 LAEVSXAVG AGHTCVTLGG DSHLAFGSIT GHAQCPDLC VIWDAHADI

151 NPLTTPSGN LHGPVSFLL RELQDKVPPI PGFSMAKPL SKSDIYIGL

201 RQDDPRAEPI LKNYDISYS MRHIDCMGK KYMEKTFDOL LGRDRPHIL

251 SFDDADPDA LAPATGTPVI GGLTYREGYV ITEEHNITGM LSAVDLYEVN

301 PYLAATSEEV KATANLADV IASCGQTRG GAHTRADTII DVLPTSTSY

351 ESDNEQVRI

11AA_SEQUENCE 1.0 STANDARD; PRT; 360 AA.

ID ARG2_XENLA

AC Q91554;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE ARGINASE, NON-HEPATIC 2 (EC 3.5.3.1).

GN ARG2.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-INTESTINE;

RX MEDLINE: 95014323.

RA PATTERSON D., SHI Y.-B.;

RT "thyroid hormone-dependent differential regulation of multiple arginase genes during amphibian metamorphosis.";

RT J Biol Chem 269:25328-25334(1994).

FT 1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED IN TISSUE REMODELING.

CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O -> L-ORNITHINE + UREA.

CC -1- COFACTOR: MANGANESE (BY SIMILARITY).

CC -1- PATHWAY: FIRST STEP IN ARGinine DEGRADATION.

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN TAIL, INTESTINE, HINDLIMB AND TRUNK REGION. STRONGEST IN TADPOLE TAIL.

CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN NEURULA (STAGE 16/17).

CC HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48. IN THE

INTESTINE, INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS (STAGES 58-64) AND IN THE HINDLIMB, EXPRESSED AT LOW LEVELS DURING METAMORPHOSIS UNTIL STAGE 66 WHEN LEVELS DRAMATICALLY INCREASE. IN THE TAIL, A CONSTANT HIGH LEVEL OF EXPRESSION IS FOUND THROUGHOUT METAMORPHOSIS.

CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).

CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: 008407; AA056892.1; -

DR HSBP; P07824; 2RLA.

DR PROSITE; PS00147; ARGINASE_1; 1.

DR PROSITE; PS00148; ARGINASE_2; 1.

DR PROSITE; PS01053; ARGINASE_3; 1.

DR PFM; PF00491; arginase; 1.

KW Urea cycle; Arginine metabolism; Hydrolase; Manganese; Multigene family.

FT METAL 122 122 MANGANESE 1 (BY SIMILARITY).

FT METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).

FT METAL 147 147 MANGANESE 2 (BY SIMILARITY).

FT METAL 149 149 MANGANESE 1 (BY SIMILARITY).

FT METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).

FT METAL 255 255 MANGANESE 2 (BY SIMILARITY).

SEQUENCE 360 AA; 39154 MW; 26CE48BC CRC32;

ARG2_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4641 ..

1 MSRSNPFRL LKQVSIIRL OKKSHSVAV IGAFSGKGR RGVHGPAA

51 IRSAGLIERL SNLCNVCDF GDHFSGVFN DELYSIVKH PRVGLACKV

101 LAEVSXAVG AGHTCVTLGG DSHLAFGSIT GHAQCPDLC VIWDAHADI

151 NPLTTPSGN LHGPVSFLL RELQDKVPPI PGFSMAKPL SKSDIYIGL

201 RQDDPRAEPI LKNYDISYS MRHIDCMGK KYMEKTFDOL LGRDRPHIL

251 SFDDADPDA LAPATGTPVI GGLTYREGYV ITEEHNITGM LSAVDLYEVN

301 PYLAATSEEV KATANLADV IASCGQTRG GAHTRADTII DVLPTSTSY

351 ESDNEQVRI

11AA_SEQUENCE 1.0 STANDARD; PRT; 360 AA.

ID ARG3_XENLA

AC Q91555;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE ARGINASE, NON-HEPATIC 3 (EC 3.5.3.1).

GN ARG3.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-INTESTINE;

RX MEDLINE: 95014323.

RA PATTERSON D., SHI Y.-B.;

RT "thyroid hormone-dependent differential regulation of multiple arginase genes during amphibian metamorphosis.";

RT J Biol Chem 269:25328-25334(1994).

FT 1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED IN TISSUE REMODELING.

CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O -> L-ORNITHINE + UREA.

```

CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN TAIL, INTESTINE, HINDLIMB AND TRUNK REGION. STRONGEST IN TADPOLE TAIL.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN EARLY TAILBUD (STAGE 23/24). HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48. IN THE INTESTINE, INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS (STAGES 58-64) AND IN THE HINDLIMB, EXPRESSED AT LOW LEVELS DURING METAMORPHOSIS UNTIL STAGE 66 WHEN LEVELS DRASTICALLY INCREASE. IN THE TAIL, A CONSTANT HIGH LEVEL OF EXPRESSION IS FOUND THROUGHOUT METAMORPHOSIS.
CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).
CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08408; AAA56893.1; ..
DR HSSP: P07824; 2RLA.
DR PROSITE: PS00147; ARGINASE_1; 1.
DR PROSITE: PS00148; ARGINASE_2; 1.
DR PROSITE: PS01053; ARGINASE_3; 1.
DR PRAM: PF00491; arginase; 1.
KM Urea Cycle: Arginine metabolism; Hydrolase; Manganese; Multigene family.
FT METAL 122 122 MANGANESE 1 (BY SIMILARITY).
FT METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 147 147 MANGANESE 2 (BY SIMILARITY).
FT METAL 149 149 MANGANESE 1 (BY SIMILARITY).
FT METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 255 255 MANGANESE 2 (BY SIMILARITY).
SQ SEQUENCE 360 AA; 39238 MW; A89683FA CRC32;

ARG3_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4724 ..

1 MSIRSFVRL LKKQVNIKL QKKCSHVAV IGAPFSKQK RRGVHSPAA
51 IRSAGLIDRL SNLCNVCDF GDHFSGQVPN DEQNSIYKH PRVYGLACKV
101 LAKEVGNVAG AGHTCVTLGG DSHLAFGSI GHACCPCDLC VIWDAHADI
151 NPLTTPSGN LHGQVPSFL RELQDKTPTI PGFSNAKCL SKSDIYIGL
201 RDDPAPQFI LKNVSIYSY MRHIDCKGR KYMETPTPOL LGRDRRIHL
251 SPIDAFDPA LAPATGTPVI GGLTYRGVY ITTEHNHGM LSAIDLEVN
301 PVLTATSEV KATANLAVDV IASCGOTRE GAHTRADITI DVLPSTSY
351 ESDNEOVRI

11AA_SEQUENCE 1.0
ID ARG3_ECOLI STANDARD; PRT; 334 AA.
AC P11446;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (EC 1.2.1.38) (AGPR) (N-ACETYL-GLUTAMATE SEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE).
GN ARG3
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89121510.

RA PARSON C., BOYEN A., COHEN G.N., GLANSDOFF N.:
" Nucleotide sequence of Escherichia coli argB and argC genes: comparison of N-acetylglutamate kinase and N-acetylglutamate-gamma-semialdehyde dehydrogenase with homologous and analogous enzymes. " ;
RL Gene 68:275-283(1988).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12; MG1655;
RX MEDLINE: 94089392.
RA BLATTER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.:
" Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 97.8 minutes. " ;
RT Nucleic Acids Res. 21:5408-5417(1993).
RL [3]
RP SEQUENCE OF 1-48 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 83143275.
RA PIETTE J., CUNIN R., BOYEN A., CHARLIER D.R.M., CRABEEL M., VAN VLIET F., GLANSDOFF N., SOUIRES C., SOUIRES C.L.:
" The regulatory region of the divergent argECB operon in Escherichia coli K-12. " ;
RT Nucleic Acids Res. 10:8031-8048(1982).
RL [4]
RP SEQUENCE OF 1-19 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 92202162.
RA MEINDEL T., SCHWITZ E., MECHULAM Y., BLANQUET S.:
" Structural and biochemical characterization of the Escherichia coli argE gene product. " ;
RT Bacteriol. 174:2323-2331(1992).
CC -1- CATALYTIC ACTIVITY: N-ACETYL-L-GLUTAMATE-5-SEMIALDEHYDE + NADP(+) + ORTHOPHOSPHATE -> N-ACETYL-5-GLUTAMYL PHOSPHATE + NADPH.
CC -1- PATHWAY: THRD STEP IN ARGININE BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER BACTERIAL ARG AND TO NAGSA DOMAIN OF FUNGAL PROTEIN ARG3/6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M21446; AAA23477.1; ..
DR EMBL: J01587; AAB59146.1; ..
DR EMBL: X55417; -; NOT ANNOTATED_CDS.
DR EMBL: U00006; AAC43064.1; ..
DR EMBL: AE000470; AAC76840.1; ..
DR PIR: J03332; RDECEP.
DR ECOGENE: BG10065; ARG3.
DR PROSITE: PS01224; ARG3.
DR PRAM: PF01118; Semialdehyde-dh; 1.
KM Arginine biosynthesis: Oxidoreductase: NADP.
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 334 AA; 35952 MW; DB98AB60 CRC32;

ARGC_ECOLI Length: 334 February 14, 2000 08:02 Type: P Check: 6342 ..

1 MLNLIYVGS GYAGALVTV VNRPHNIT ALTVASQND AGKILSDLP
51 QLKGIYDPL QPMDSISES PGVNVFLAT ABEVSHDLP QLEAGCVF
101 DLGAFRND ATFEKYTF THQYPELLEQ AAYGLAEMWG NKLEANLIA
151 VPGCTPRAQ LAKPLIDAD LLDLNPYVI NATSGVSGAG RKAALNSFC
201 EYSLQPIGVF TRHOPFELT HLGADVIFTP HLGNEFRGLI ETTCRLKSG
251 VTAQVAOVL QQAYAHKPLV RLYDKGPAL KNYVGLPFCD IGFAVGEHL

```



```

1 MOUNTAIN TITMICE FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WPKTRKYS PHSLPQX
11AA_SEQUENCE 1.0 STANDARD; PRT; 68 AA.
ID ATP8_PONPY
AC P30049;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A6L).
GN MTAP8 OR ATP8.
OS Pongo pygmaeus pygmaeus (Bornean orangutan).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Pongo.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATES ANNA AND DENNIS;
RX MEDLINE: 97032590.
RA XU X., ARNASON U.;
RT "The mitochondrial DNA molecule of Sumatran orangutan and a molecular
RT proposal for two (Bornean and Sumatran) species of orangutan.";
RL 1. Mol. Evol. 43:431-437(1996).
CC 1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC 1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC 1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X97715; CAN66301.1;
DR EMBL: X97711; CAN66297.1;
DR PFM: PF00895; ATP-syn-8; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ TRANSMEM 8 24 POTENTIAL.
ATP8_PONPY Length: 68 February 14, 2000 08:02 Type: P Check: 3460
1 MOUNTAIN TITMICE FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WPKTRKYS PHSLPQX
11AA_SEQUENCE 1.0 STANDARD; PRT; 68 AA.
ID ATP8_PONPY
AC P30584;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A6L).
GN MTAP8 OR ATP8.
OS Pongo pygmaeus (Orangutan).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Pongo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 95132634.
RA HORAI S., HAYASAKA K., KONDO R., TSUGANE K., TAKARATA N.;
RT "Recent African origin of modern humans revealed by complete sequences
RT of hominoid mitochondrial DNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
CC 1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC 1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC 1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D38115; BAA07309.1;
DR PFM: PF00895; ATP-syn-8; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ TRANSMEM 8 24 POTENTIAL.
ATP8_PONPY Length: 68 February 14, 2000 08:02 Type: P Check: 3550
1 MOUNTAIN TITMICE FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WPKTRKYS PHSLPQX
11AA_SEQUENCE 1.0 STANDARD; PRT; 168 AA.
ID ATPD_HUMAN
AC P30049;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR (EC 3.6.1.34).
GN ATP5D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92182007.
RA JORDAN E.M., BREEN G.A.M.;
RT Molecular cloning of an import precursor of the delta-subunit of the
RT human mitochondrial ATP synthase complex.";
RL Biochim. Biophys. Acta 1130:123-126(1992).
CC 1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC 1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC SEEMS TO HAVE NINE SUBUNITS: A, B, C, D, E, F, G, F6 AND 8 (OR
CC A6L).
CC 1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC 1- SIMILARITY: BELONGS TO THE ATPASE SUBUNIT CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X63422; CAA45016.1; -
 CC EMBL: X63423; CAA45017.1; -
 CC EMBL: AC004221; AAC04304.1; -
 DR PIR: S22348; S22348.
 DR SWISS-2DPAGE: P30049; HUMAN.
 DR MIM: 603150; -
 DR PFAM: PF00401; ATP-synt-DE; 1.
 KW ATP synthetase; CF(1); Hydrogen ion transport; Hydrolase;
 KW Mitochondrion; Transit peptide.
 FT TRANSLIT 1 22 MITOCHONDRION.
 FT CHAIN 23 168 ATP SYNTHASE DELTA CHAIN.
 SQ SEQUENCE 168 AA; 17490 MW; 0182AE71 CRC32;

ATPD_HUMAN Length: 168 February 14, 2000 08:02 Type: P Check: 1686 ..

1 MLEPALLRRP GIGRLVRRAR AYAEAAAP AASGPNOMSF TFASPTGVFF
 51 NGANVQGVY PLTGAFGIL AAHYPLQVL RGLVYVAE DGTSTKYFVS
 101 SGIAVNADS SVQLAEAV TLMDLGAA KANLEKQAE LVGTADATR
 151 AEQIRIEAN EALYKALE

11AA_SEQUENCE 1.0 STANDARD: PRT: 127 AA.
 ID ATP2_BACP3
 AC P09354;
 DT 01-MAR-1989 (Rel. 10; Created)
 DT 01-NOV-1997 (Rel. 35; Last annotation update)
 DE ATP SYNTHASE PROTEIN I.
 OS Bacillus p33 (Thermophilic bacterium PS-3).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86163679.
 RA OTAMA S., YOHDA M., ISHIZUKA M., HIRATA H., HANAMOTO T.,
 RT "Sequence and over-expression of subunits of adenosine triphosphate
 RT synthase in thermophilic bacterium PS-3".
 RL Biochim. Biophys. Acta 933:141-155(1988).
 CC -1- FUNCTION: A POSSIBLE FUNCTION FOR THIS PROTEIN IS TO GUIDE THE
 CC ASSEMBLY OF THE MEMBRANE SECTOR OF THE ATPASE ENZYME COMPLEX.
 CC -1- SIMILARITY: TO THE CORRESPONDING SUBUNIT IN OTHER BACTERIA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X07804; CAA30647.1; -
 DR PIR: S01397; S01397.
 KW Hydrogen ion transport; CF(0); Transmembrane.
 SQ SEQUENCE 127 AA; 14595 MW; 5999D881 CRC32;

ATP2_BACP3 Length: 127 February 14, 2000 08:02 Type: P Check: 4392 ..

1 MGNLQANFWR QVRILYLLA IYTLGFTPT YKTVFSLIL GTSISLAWW
 51 NLTKIKKFG QAVAAKKVR TLGLSLRAL AALAAVIVL YPQYHIYPT
 101 VLGIMTSYIV IIDFFHKK KNDKILQA

11AA_SEQUENCE 1.0 STANDARD: PRT: 398 AA.
 ID BMAP_XENLA

P25703;
 DT 01-MAY-1992 (Rel. 22; Created)
 DT 01-MAY-1992 (Rel. 22; Last sequence update)
 DT 01-JUL-1993 (Rel. 26; Last annotation update)
 DE BONE MORPHOGENETIC PROTEIN 2-1 PRECURSOR (BMP-2-1).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Piplidae; Xenopodinae;
 CC Xenopus.
 CC [1]
 CC SEQUENCE FROM N.A.
 RX MEDLINE: 91274367.
 RA PLESSEY S., KOESTER M., KNOECHER W.;
 RT "cDNA sequence of Xenopus laevis bone morphogenetic protein 2
 RT (BMP-2)".
 RL Biochim. Biophys. Acta 1089:280-282(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92378616.
 RA NISHIMATSU S., SUZUKI A., SHODA A., MURAKAMI K., UENO N.;
 RT "Genes for bone morphogenetic proteins are differentially transcribed
 RT in early amphibian embryos".
 RL Biochem. Biophys. Res. Commun. 186:1487-1495(1992).
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X55031; CAA38850.1; -
 DR EMBL: X63424; CAA45018.1; -
 DR PIR: S16244; S16244.
 DR PIR: JH0687; JH0687.
 DR HSSP: P18075; 1BMP.
 DR PROSITE: PS00250; TGF-BETA: 1.
 DR PFAM: PF00019; TGF-beta: 1.
 KW PFAM: PF00688; TGF-beta; TGF-beta; 1.
 KM Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
 FT SIGNAL 1 284
 FT CHAIN 1 284
 FT PROPE 285 398
 FT DISULFD 298 363
 FT DISULFD 327 395
 FT DISULFD 331 397
 FT DISULFD 362 362
 FT CARBOHYD 137 137
 FT CARBOHYD 202 202
 FT CARBOHYD 340 340
 FT CONFLICT 7 7
 FT CONFLICT 16 16
 FT CONFLICT 233 233
 SQ SEQUENCE 398 AA; 45575 MW; 6143F996 CRC32;

BMAP_XENLA Length: 398 February 14, 2000 08:02 Type: P Check: 9714 ..

1 MVAGISLILL LLFYQVLLSG CTGLPEEGK RRYTESGRSS PQOSORVLNQ
 51 FELRLSMFG LKRPTPGKN VVIPYMLD YHLILAOLAA DEGTSMDFQ
 101 MERAASRANT VRSFHESM EEIPESREKT IQRFENLSS IPNEELVISA
 151 ELRIPEQVQ EPFESDSSKL HRINIDYK PAALASQPV VRLDTRLVH
 201 HNSKKESTF VTPALARMIA HKOPNGFV EVNHLNDKN VPKHVRISR
 251 SLTPDKNWP QIRPLVTF S HDGKGHALK RQKQARRK RKLKSCSR

```

301 HPLIYDFSDV GNMWIVAPP GYHAFYCHGE CPEPLADHLN STNHAIVOTL
351 VNSVNTNIPK ACQVPELISA ISMLYIDENE KYVLKNYODM VVEGCGCR

!!AA_SEQUENCE 1.0
ID BMBP_XENLA STANDARD: PRT: 398 AA.
AC P30884;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE BONE MORPHOGENETIC PROTEIN 2-II PRECURSOR (BMP-2-II).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92378616.
RA NISHIMATSU S., SUZUKI A., SHODA A., MURAKAMI K., UENO N.;
RT "Genes for bone morphogenetic proteins are differentially transcribed
in early amphibian embryos."
RL Biochem Biophys Res Commun. 186:1487-1495(1992).
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
CC -1- SUBUNIT: HOMODIMER. DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63425; CAA45019.1; -
DR F01R; JH0688; JH0688.
DR HSSP: P18075; IBM.
DR PROSITE: PS00250; TGF_BETA; 1.
DR PFAM: PF00019; TGF-beta; 1.
KW Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
FT SIGNAL 1 ?
FT PROPEP 284 ?
FT CHAIN 285 398
FT DISULFID 298 363
FT DISULFID 327 395
FT DISULFID 331 397
FT DISULFID 362 362
FT CARBOHYD 137 137
FT CARBOHYD 202 202
FT CARBOHYD 237 237
FT CARBOHYD 340 340
SQ SEQUENCE 398 AA; 45616 MW; E7DIDBA CRC32;

BMBP_XENLA Length: 398 February 14, 2000 08:02 Type: P Check: 1620
1 MAGHSLL LQFOILLG CTGLVPEEGK RRYSESTRSS PQOSQOVLDO
51 FERRLLNMG LKRPTPEKN VIPPYWLDI YHLHSAQLAD DQGSSEVDYH
101 MERAASRANT VRSPHHEESM EEIPESGERT IORFFNLSS IPDELYTSS
151 ELRIREBOVQ EPEFTDQSKL HRINIYDIK PAAASRGPV VRLDTRLIH
201 HNESWESFD VTPAIRWIA HKOPNHGFV EYTHLDNDTN VPKRWISR
251 SLTLDRGHP RIRPLVTS HGKGHALK ROKROARKQ KRKLKSCCR
301 HPLVDFSDV GNMWIVAPP GYHAFYCHGE CPEPLADHLN STNHAIVOTL
351 VNSVNTNIPK ACQVPELISA ISMLYIDENE KYVLKNYODM VVEGCGCR

!!AA_SEQUENCE 1.0

```

```

ID BPHB_PSEPS STANDARD: PRT: 275 AA.
AC P08654;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1991 (Rel. 34, Last annotation update)
DE BIPHENYL-2,3-DIHYDRO-2,3-DIOL DEHYDROGENASE (EC 1.3.1.-) (BIPHENYL-
DE CIS-DIOL DEHYDROGENASE) (2,3-DIHYDROXY-4-PHENYLHEXA-4,6-DIENE
DE DEHYDROGENASE).
GN BPHB.
OS Pseudomonas pseudoalcaligenes.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-KF707;
RX MEDLINE: 87083404.
RA FURUKAWA K., ARIMURA N., MIYAZAKI T.;
RT "Nucleotide sequence of the 2,3-dihydroxybiphenyl dioxygenase gene of
RT Pseudomonas pseudoalcaligenes."
RL J. Bacteriol. 169:427-429(1987).
CC -1- PATHWAY: DEGRADATION OF BIPHENYLS AND POLYCHLOROBIPHENYLS (PCB) TO
CC BENZOIC ACID AND CHLOROBENZOIC ACIDS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15333; AAA25752.1; ALT-INIT.
DR HSSP: P4722; IBD.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
FT ACT_SITE 9 33
FT BIND 153 153
FT SITE 153 153
FT SEQUENCE 275 AA; 28726 MW; EA999DF7 CRC32;

BPHB_PSEPS Length: 275 February 14, 2000 08:02 Type: P Check: 1527
1 MKLGEAVLI TGGASGLGRA LVDFEVAEK VAVLDKSAER LALEETDLGD
51 NVLGIQDVR SLEDQKQAS RCVARFGKID TLIPNAGTWD YSTALVDLPE
101 ESDDAFDEV FHINVKGYIH AYALPALVA SGNVIFITIS NAGRYPNGGG
151 PLYTAKQAI VGLVRELAFE LARYVANGV GGGGNSDMR GPSSLGSGSK
201 AISTVPLADM LKSVLPIDRM PEVEEYTGAY VFPAIRGDA PASGALVNYD
251 GGLGVGCFPS GAGNDLLDQ LNIHP

!!AA_SEQUENCE 1.0
ID BR3A_HUMAN STANDARD: PRT: 423 AA.
AC G01851; Q15318; Q14865.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1991 (Rel. 35, Last sequence update)
DT 01-NOV-1991 (Rel. 35, Last annotation update)
DE BRAIN-SPECIFIC HOMEOBOX/POU DOMAIN PROTEIN 3A (BRN-3A) (OCT-11)
DE (HOMEOBOX/POU DOMAIN PROTEIN RDC-1).
GN FOD4F1 OR BRN3A OR RDC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX XIANG M., ZHOU L.-J., MACKE J.P., EDDY R.L., SHOMS T.B., NATHANS J.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

```

[2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 94052142.
 RA BHARGAVA A.K., LI Z., WEISSMAN S.M.;
 RT "differential expression of four members of the pou family of
 RT proteins in activated and phorbol 12-myristate 13-acetate-treated
 RT Burkitt T cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10260-10264(1993).
 RN
 RP SEQUENCE OF 85-423 FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 93027214.
 RA COLLUM R.G., FISHER P.E., DATTA M., MELLIS S., THIELE C.,
 RA HUEBNER K., CROCE C.M., ISRAEL M.A., THEIL T., MOROY T., DEPINHO R.,
 RA ALT F.W.;
 RT "A novel Pou homeodomain gene specifically expressed in cells of the
 RT developing mammalian nervous system";
 RL Nucleic Acids Res. 20:4919-4925(1992).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN
 CC THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF
 CC NEURONAL LINEAGES. MAY PLAY A ROLE IN DETERMINING OR MAINTAINING
 CC THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: BRAIN. SEEMS TO BE SPECIFIC TO THE RETINA.
 CC PRESENT IN THE DEVELOPING BRAIN, SPINAL CORD AND EYE.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS EARLY IN EMBRYOGENESIS
 CC (DAY 13.5) AND IS UNDETECTABLE 14 DAYS AFTER BIRTH.
 CC -1- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
 CC TO CLASS-4 POU.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U10063; AAA57161.1; -
 DR EMBL: U10062; AAA57161.1; JOINED.
 DR EMBL: L20433; AAA65605.1; -
 DR EMBL: X64624; CAA45907.1; -
 DR HSSP: P10037; 1A07.
 DR MIM: 601632; -
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00035; POU_1; 1.
 DR PROSITE: PS00465; POU_2; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PFAM: PF00046; homeobox; 1.
 DR PFAM: PF00157; pou; 1.
 DR Transcription regulation: Nuclear protein; DNA-binding; Homeobox;
 KW Developmental protein.
 KM
 FT DOMAIN 57
 FT DOMAIN 106
 FT DOMAIN 130
 FT DOMAIN 189
 FT DOMAIN 260
 FT DOMAIN 267
 FT DOMAIN 341
 FT DNA_BIND 359
 FT CONFLICT 91
 FT CONFLICT 92
 FT CONFLICT 99
 FT CONFLICT 133
 FT CONFLICT 135
 FT CONFLICT 137
 FT CONFLICT 139
 FT CONFLICT 142
 FT CONFLICT 149
 FT CONFLICT 155
 FT CONFLICT 156
 FT CONFLICT 174
 FT CONFLICT 175
 FT CONFLICT 189
 FT CONFLICT 240
 FT CONFLICT 240
 FT CONFLICT 252
 FT CONFLICT 256
 FT CONFLICT 276
 FT CONFLICT 343
 FT SEQUENCE 423 AA; 42939 MW; 0DC8F2BD CRC32;

BRAL_HUMAN Length: 423 February 14, 2000 08:02 Type: P Check: 2610 ..
 1 MMSNSKOPH FAMHPTLPEH KPSLSHSSE AIRACLPTE PLOSNILFASL
 51 DELLARAELA LAAVDIAVSQ GKSHPFKPDA TYHTMSVPC TSTVPLAH
 101 HHHHHHHQA LEPGDLIDHT SSPSLALWAG AGGAGGAGAA AGGGAHDGP
 151 GGGGGPGGGG GPGGGPGGGG GGGGPGGGG GPGGGLLGS AHPHHHMSL
 201 GHLSPAAA AMMPGSLPH POLVAAAHH GAAAAAAGAA AGVAAAAGAA
 251 AAAGAAAGAA STCDSDTDR ELEAFAPFK QRRKLGTV AVGSALANL
 301 KIPGVSLSQ STICREBSLT LSHNMWLLK PIIQWILEA EAQDEKKNK
 351 PELFGGERK RKRSTIAPE KRSLAYFAY QPRPSSEKIA AIAERLCLK
 401 NVVRVFCNQ ROKOKRKES ATY
 !!AA_SEQUENCE 1.0
 ID BRAL_BRAFL STANDARD; PRT; 448 AA.
 AC Q17134;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BRACHYURY PROTEIN HOMOLOG 1 (AMBRA-1).
 GN BRA-1.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LARVA
 RX MEDLINE: 96125169.
 RA HOLLAND P.W.H., KOSCHORZ B., HOLLAND L.Z., HERRMANN B.G.;
 RA "Conservation of Brachyury (T) genes in amphioxus and vertebrates:
 RT developmental and evolutionary implications";
 RL Development 121:4283-4291(1995).
 CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
 CC REQUIRED FOR MESODERM FORMATION AND DIFFERENTIATION
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN THE 10 H GASTRULA AROUND
 CC THE BLASTOPORE. IN LATER STAGES OF GASTRULATION AND EARLY
 CC NEURALIZATION, EXPRESSED IN POSTERIOR MESODERM AND THE NOTOCHORD
 CC WHERE EXPRESSION CONTINUES AT LEAST UNTIL SWIMMING LARVAL STAGE.
 CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X91503; CAA62999.1; -
 DR HSSP: F24781; 1XBR.
 DR PROSITE: PS01264; TBOX_2; 1.
 DR PROSITE: PS01283; TBOX_1; 1.
 DR PFAM: PF00907; T-box; 1.
 KW Developmental protein; Transcription regulation; DNA-binding;
 KW Nuclear protein; Multigene family.
 FT DNA_BIND 54
 FT SEQUENCE 448 AA; 49253 MW; 5948F2D3 CRC32;
 BRAL_BRAFL Length: 448 February 14, 2000 08:02 Type: P Check: 7365 ..
 1 MSAEMKOP TAASPOFSV SHLSAASE ISASEKGP TENDLITILE
 51 EKLMDKRNA LTNEMIVTKN GRMPEVLKV NVSGLDENAM YSLDLFTAA

101 DNRKMYNG EWPGRKPEP SVPSVYIHP DSPNFGAMM KSPVSFSVK
 151 LTNKLNKGGG IMLNSLHKE PRHIIKVG PNQNMVSTH TPETOTIAY
 201 TAYONEIITA LKIKNPAP AFIDAKERSD GKDGHEDDO QPQSQSGM
 251 FLPGTPICP PPNHQFAPS LGPSHGCDR YSTLANHSA PYHPYORSS
 301 PPNYGHDTA ASLPMPTHD NMGSLPVSTH NMLSNMNH TTTSTHAQY
 351 NLMVSNNML TPTTHAQTH SGTMGGLPH QFLRTTAPAP YHSIPICTVP
 401 TTASSPVYH DSHEVSTDS GYGHSTTPA POTRITSNM SPMTPLSL

11AA_SEQUENCE 1.0
 ID BRA2_BRAFL STANDARD: PRT: 440 AA.
 AC P80492;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BRACHYURY PROTEIN HOMOLOG 2 (AMBR-2).
 GN BRA-2.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96125169.
 RA HOLLAND P.W.H., KOSCHORZ B., HOLLAND L.Z., HERRMANN B.G.;
 RT "Conservation of Brachyury (T) genes in amphioxus and vertebrates:
 development and evolutionary implications."
 RL Development 121:4283-4291(1995).
 CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
 REQUIRED FOR MESODERM FORMATION AND DIFFERENTIATION
 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: MESODERM AND NOTOCHORD.
 CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
 DR HSSP: P24781; 1XBR.
 DR PROSITE: PS01264; TBOX_2; 1.
 DR PFAM: PF00907; T-box; 1.
 DR Developmental protein; Transcription regulation; DNA-binding;
 KW Nuclear protein; Multigene family.
 FT DNA BIND 44
 FT SEQUENCE 440 AA; 48469 MW; 4828EB28 CRC32;

BRA2_BRAFL Length: 440 February 14, 2000 08:02 Type: P Check: 391

1 MKOTPOFSV SHLSAVERSE ISAGSEKGP TERDKATLG EKPLMEKFS
 51 LTNEMIVTYS GRMFPVLKV NVGLDPMAM YSLDLPTFAA DNRKMYNG
 101 EWPGRKPEP SVPSVYIHP DSPNFGAMM KSPVSFSVK LTNKLNKGGG
 151 QIMLSLHKE PRHIIKVG GPDNORTLS HTFAETOFIA VTAYONEILT
 201 ALKIKNPAP AFIDAKERN DTKSGHDLT DOOPFOSLG GWPLPGTPI
 251 CPPNHOFA PSIGLPSHC DRYSTLNRH SAPYHPYOR SSPPTNGHD
 301 TAASLPMPT HDNMGSLPV THNMLMSAM PHTTSTHAQ YPNLMVSNN
 351 NLPTTHAQT HMSGTGTGL PHQFLRTAP APYHSIPCT VPTASSPV
 401 YHSHSVST DSGYGHSTP PAROTRITS NMSPTNMSM

11AA_SEQUENCE 1.0
 ID BSR_BACCE STANDARD: PRT: 140 AA.
 AC P3367;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)

01-NOV-1995 (Rel. 32, Last annotation update)
 DE BLASTICIDIN-S DEAMINASE (EC 3.5.4.23).
 GN BSR.
 OS Bacillus cereus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-46.
 RC STRAIN-K5-S1;
 RX MEDLINE: 92144117.
 RA KOBAYASHI K., KAKIKURA T., TANAKA T., YAMAGUCHI I., ENDO T.;
 RT "Nucleotide sequence of the bsr gene and N-terminal amino acid
 sequence of blastcidin S deaminase from blastcidin S resistant
 Escherichia coli T121."
 RL Agric. Biol. Chem. 55:3155-3157(1991).

CC -1- FUNCTION: CATALYZES THE DEAMINATION OF THE CYTOSINE MOIETY OF THE
 CC -1- ANTIBIOTIC BLASTICIDIN S, CYTOMYCIN AND ACETYLBASTICIDIN S.
 CC -1- CATALYTIC ACTIVITY: BLASTICIDIN S + H(2)O =
 CC -1- DEAMINOHYDROXYBLASTICIDIN S + NH(3).
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
 CC FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----

DR EMBL: S81409; AAC60404.1; -
 DR PIR: J50609; J50608
 DR PROSITE: PS00905; CYT_DCMP_DEAMINASES; 1.
 DR PFAM: PF00385; dCMP_cyt_deam; 1.
 KW Hydrolyase; Zinc.
 FT METAL 59
 FT METAL 59
 FT METAL 100
 FT METAL 103
 FT METAL 103
 FT ZINC (BY SIMILARITY).
 FT ZINC (BY SIMILARITY).
 FT ZINC (BY SIMILARITY).
 SO SEQUENCE 140 AA; 15573 MW; 12D8CACC CRC32;

BSR_BACCE Length: 140 February 14, 2000 08:02 Type: P Check: 8425

1 MKTNISOOD LELVEATEK ITMLEDNKH HVGCAIRTKT GEISAVHIE
 51 AYIGRVYCA EALIGSAVS NGQKDEPTV AVRHPSDEV DRSIRVSPC
 101 GMRCLISDY APDCFLIEM NGKLKTTIE ELPLKVTEN

11AA_SEQUENCE 1.0
 ID CAD1_DICDI STANDARD: PRT: 213 AA.
 AC P54637;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CALCIUM-DEPENDENT CELL ADHESION MOLECULE-1 (DCCAD-1) (GP24).
 GN CADA.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Dictyostelida; Dictyostelium.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-62 AND 118-133.
 RX MEDLINE: 96279194.
 RA WONG E.F.S., BRAR S.K., SESAKI H., YANG C., SIU C.-H.;
 RT "Molecular cloning and characterization of Dccad-1, a Ca²⁺-dependent
 cell-cell adhesion molecule, in dictyostelium discoideum."
 RL J. Biol. Chem. 271:16399-16408(1996).
 CC -1- FUNCTION: INVOLVED IN CALCIUM-DEPENDENT CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE ECTO-SURFACE OF THE
 CC PLASMA MEMBRANE. MAY BE TRANSPORTED TO THE PLASMA MEMBRANE VIA
 CC CONTRACTILE VACUOLAE AND ITS CELL SURFACE ASSOCIATION MAY BE
 CC MEDIATED BY AN INTEGRAL MEMBRANE PROTEIN
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED SOON AFTER THE INITIATION OF
 CC DEVELOPMENT.


```

CC -1- DOMAIN: CELL BINDING ACTIVITY IS DEPENDENT ON THE N-TERMINAL
CC SEGMENT AND NOT ON THE PUTATIVE CA(2+)-BINDING SITE IN THE C-
CC TERMINUS.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49650; AAC47135.1; -
DR DICTDB: DD02037; CADA.
KW Cell Adhesion; Calcium; Repeat.
FT DOMAIN 1 184 4 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 97 1.
FT REPEAT 49 97 2.
FT REPEAT 98 146 3.
FT REPEAT 147 194 4.
FT SEQUENCE 213 AA; 23926 MW; DB554D11 CRC32;
CADI_DICDI Length: 213 February 14, 2000 08:02 Type: P Check: 3813

1 MSVDANKVF FFGKNCGTGS FEYKGETVR FNNGDKNDK FMSCLVGSNV
51 RCNIWEHNEI DTPPGKFOE LAQSTNNDL TSINGISKFO VLPAGAFOMAV
101 DVKIVNRYNS TAGSEYMTIT PYQYDVACK DGDDEYOLPI PKLPPDSEI
151 VSHLTVRQTH TPYDYVNGS VYFKYSPFG QYTVIKDET FPKNMTVTD
201 DNTSEIFNLN SEK

11AA_SEQUENCE 1.0 STANDARD: PRT; 294 AA.
ID CANG_MOUSE
AC P49070;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CALCIUM-SIGNAL MODULATING CYCLOPHILIN LIGAND (CAML).
GN CAML6 OR CAML.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B6/CBA; TISSUE-THYMUS;
RX MEDLINE; 96011657.
RA KIM H.S., MORALES V.M., DASS C., ENCINAS J., TETTEL M.,
RA BLUMBERG R.S.;
RT "Cloning of the gene encoding the mouse homologue of the human
RT calcium signal-modulating ligand."
RL Gene 163:323-324(1995).
CC -1- FUNCTION: LIKE THE TCR/CD3 COMPLEX INTERACTION, BINDS TO CYCLOPHILIN B.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U1960; AAA87004.1; -
DR MGD: MGI:104728; CAML.
KW Transmembrane
KM TRANSMEM 188 208
FT TRANSMEM 239 255 POTENTIAL.
FT SEQUENCE 294 AA; 32514 MW; 6BD303C5 CRC32;

```

```

CANG_MOUSE Length: 294 February 14, 2000 08:02 Type: P Check: 283

1 MEVPATATG GERPATPSGL SASORAEAR RRLKLNSEQ RINRWGFR
51 PGSGSEENG TSKRQPSDK LNSLSPSVS KRVLGDSVD GGGADLGCV
101 AFRGQLOLD KUDSPKAP CSKDKDAELR QRTRODLTAD PAORASHHL
151 EQLSREER MTKRKQLSE KPSQEDGSTA EEPDFRFR LVGCALLALG
201 VRAFVCKYIS IPAPLTQL AYMGVKKYF KGEKVKTV LTAALLSGI
251 PAEAINRMD TYSKGEVFT DLGVYFTFI FCHELDYWG PEVP

11AA_SEQUENCE 1.0 STANDARD: PRT; 136 AA.
ID CDD_BACSU
AC P19079;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYTIDINE DEAMINASE (EC 3.5.4.5) (CYTIDINE AMINOHYDROLASE) (CDA).
GN CDD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 89313687.
RA SONG B.-H., NEHARD J.;
RT "Chromosomal location, cloning and nucleotide sequence of the
RT Bacillus subtilis cdd gene encoding cytidine/deoxycytidine
RT deaminase."
RT Mol. gen. genet. 216:462-468(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH62;
RA KOBAYASHI Y., KIZONO M., MASUDA S., TAKEKAWA K., HOSONO S.,
RA SAITO T., TAKEUCHI M.;
RT Submitted (May-1996) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN-ED40;
RA KIM K., HWANG S., SUH J., SONG B.-H., HONG S., KIM J.;
RT "Nucleotide sequence upstream of the cdd locus in Bacillus subtilis."
RT J. Microbiol. Biotechnol. 5:0-0(1995).
CC -1- FUNCTION: THIS ENZYME SCAVENGE EXOGENOUS AND ENDOGENOUS CYTIDINE
CC AND 2'-DEOXYCYTIDINE FOR UMP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: CYTIDINE + H(2)O = URIDINE + NH(3).
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18532; AAB59993.1; -
DR EMBL: X17432; NOT ANNOTATED_CDS.
DR EMBL: D84432; BAA1248.1; -
DR EMBL: U29117; AAB7045.1; -
DR EMBL: K02114; AAB03347.1; -
DR EMBL: Z89116; CAB14459.1; -
DR EMBL: Z89117; CAB14472.1; -
DR PIR: JEU022; JEU022.
DR HSSP: P13652; IAF2.
DR SUBTILIST; BG10477; CDD.

```

DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; 1.
 DR PFAM: PF00383; dCMP_cyt_deam; 1.
 KW Hydrolyase; Zinc.
 FT METAL 53 53 ZINC (BY SIMILARITY).
 FT METAL 86 86 ZINC (BY SIMILARITY).
 FT METAL 89 89 ZINC (BY SIMILARITY).
 SO SEQUENCE 136 AA; 14854 MW; 07621111 CRC32;

CDD_BACSU Length: 136 February 14, 2000 08:02 Type: P Check: 9048 ..

1 MNRQELITPA LKARDMAVAP YSKFOYGAL LTKDKYVRG CNIENAYSM

51 CMCARTALF KAYSEGDEF QMLAVADIP GPVSPCGACR QVISELCTKD

101 VIVVLTNLOG QIKETVVEL LPGAESSEDL HDERKL

11AA_SEQUENCE 1.0 STANDARD: PRT: 133 AA.

ID CDD_MYCPI STANDARD: PRT: 133 AA.

AC P47718:

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE CYTIDINE DEAMINASE (EC 3.5.4.5) (CYTIDINE AMINOHYDROLASE) (CDA).

GN CDD.

OS Mycoplasma plium.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BER:

RY MEDLINE: 9352438.

RT THAM T.N., PERRET S., KOVACIC R., MONTAGNER L., BLANCHARD A.;

RT "Identification of Mycoplasma plium genes involved in the salvage

RT pathways for nucleosides".

RL J. Bacteriol. 175:5281-5285(1993).

CC -1- FUNCTION: THIS ENZYME CATALYZES EXOGENOUS AND ENDOGENOUS CYTIDINE

CC AND 2-DEOXYCYTIDINE FOR UMP SYNTHESIS (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: CYTIDINE + H(2)O -> URIDINE + NH(3).

CC -1- COFACTOR: TIGHTLY BIND A SINGLE ATOM OF ZINC (BY SIMILARITY).

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PHOSPHATIDATE CYTIDYLTRANSFERASE (EC 2.7.7.41) (CDP-DIGLYCERIDE
 DE SYNTHASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE) (CDP-DIACYLGlycerol
 DE SYNTHASE) (CDS) (CPT:PHOSPHATIDATE CYTIDYLTRANSFERASE) (CDP-DAG
 DE SYNTHASE).
 GN CDS OR CDS OR H10919.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;
 OC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20:
 RX MEDLINE: 9530630.
 RA FLEISCHMAN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAUGE A.R., BUTT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZGUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORHAGEN N.S.M.,
 RA GENEH C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RT Science 269:496-512(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-33:
 RC LOOSMORE S.M., YANG Y., COLEMAN D.C., SHORTREED J.M., ENGLAND D.M.,
 RA KLEIN M.H.;
 RA Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: CTP + PHOSPHATIDATE -> PYROPHOSPHATE +
 CC CDP-DIACYLGLYCEROL.
 CC -1- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CDS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U32773; AAC22577.1; -

DR EMBL: U60831; AAB61967.1; -

DR EMBL: U60832; AAB61972.1; -

DR TIGR: H10919; -

DR PROSITE: PS01315; CDS; 1.

DR PFAM: PF01148; Cytidylyltransf. 1.

KW Transferrase; Nucleotidyltransferase; Phospholipid biosynthesis;

KW Transmembrane; Inner membrane

FT TRANSMEM 10 30 POTENTIAL.

FT TRANSMEM 52 72 POTENTIAL.

FT TRANSMEM 89 109 POTENTIAL.

FT TRANSMEM 118 138 POTENTIAL.

FT TRANSMEM 152 172 POTENTIAL.

FT TRANSMEM 192 212 POTENTIAL.

FT TRANSMEM 223 243 POTENTIAL.

FT VARIANT 16 16 V -> F (IN STRAIN EGMAN).

FT VARIANT 59 59 T -> A (IN STRAIN 33).

FT VARIANT 81 81 R -> H (IN STRAIN EGMAN).

SO SEQUENCE 288 AA; 32247 MW; 652F5561 CRC32;

CDS_HAEIN Length: 288 February 14, 2000 08:02 Type: P Check: 8973 ..

1 MKQRYLSAI VLIAAVICAL FLTPPYFAL ALGAVAIIGI WENQGFARLK

51 QPLIRFVIT FLVGIFILML YTEGNTLDAG RVEEDHDL LNNVSMWGL

101 ALLVYISPK SAKFWSKNPL LQLLPAPSTL IPFAGVLR LREHYTHDPY

151 HGLLELLYVE ILYWADSGA YFSGRAFEGR KLAPEVSPCK SMEGIVGGLI
 201 TALVALAFETI HESNNLVGD RNINGFILLS VATVAISVLG DLTESEMFRE
 251 SGVNDSSQLI PGHGVLDRI DGLTAAVFF SYFFFFVL
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 225 AA.
 ID CDSN_PIG
 AC 019084;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CORNDESMOSIN (S PROTEIN) (FRAGMENT).
 GN CDSN.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PERIPHERAL BLOOD;
 RA VALTEN F.W., ROGEL-GALLARD C., RENARD C., PONTAROTTI P.,
 RA TAZI-AHNINI R., VAIMAN M., CHARON P.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED; FOUND IN CORNDESMOSOMES; THE
 CC -!- INTERCELLULAR STRUCTURES THAT ARE INVOLVED IN DESQUAMATION (BY
 CC SIMILARITY).
 CC -!- TISSUE SPECIFICITY: EXCLUSIVELY EXPRESSED IN SKIN (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; 297402; CAB10690.1; -
 KW Glycoprotein.
 FT NON_TER 1
 FT DOMAIN 79 85 POLY-SER.
 FT DOMAIN 129 140 POLY-SER.
 FT DOMAIN 159 168 POLY-SER.
 FT DOMAIN 179 184 POLY-SER.
 FT NON_TER 225
 SO SEQUENCE 225 AA; 21644 MW; FE1295F0 CRC32;
 CDSN_PIG Length: 225 February 14, 2000 08:02 Type: P Check: 9217 ..
 1 GKPCPITSV DKFGSYEVVG GSSDYLVPV MTYGGKIYP VGYFKNDPV
 51 KSGPAPSPA AGPISSEKY FSSNPIPSH SSSSSNIYQS GASSAIVFOP
 101 VGGGQVPCG VSSGSKGPC SLGSGVHSS SSISSSSGSS FHPGTVSOG
 151 PCSPGTGSF SSSSSKSGG KIILPCGSK SSSSGHPCIS VSSSTLSGSP
 201 DSGPDPSPA GAKPGSGSS GKIPC
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 415 AA.
 ID CG21_SCHPO
 AC P2485;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE G2/MITOTIC-SPECIFIC CYCLIN C1G1.
 GN C1G1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.

EX MEDLINE; 91300545.
 RA BUENO A., RICHARDSON H.E., REED S.I., RUSSELL P.;
 RL "A fission yeast B-type cyclin functioning early in the cell cycle."
 RL 66:149-159(1991).
 CC -!- FUNCTION: REQUIRED FOR EFFICIENT PASSAGE OF THE G1/S TRANSITION.
 CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; M68881; AAA35288.1; -
 DR PIR; A40316; A40316.
 DR HSSP; P20248; JUST.
 DR PROSITE; PS00292; CYCLINS; 1.
 DR PFAM; PF00134; cyclin; 1.
 KW Cyclin; Cell cycle; Cell division; Mitosis.
 SO SEQUENCE 415 AA; 47827 MW; ED01A48C CRC32;
 CG21_SCHPO Length: 415 February 14, 2000 08:02 Type: P Check: 7358 ..
 1 MDVSTOTRRA TYFDENQLQ KHIIVYKKS HIKLNTGVA PKKAVDNIQO
 51 QDEPTLIEGN NESSISSSTG DFEEDFAVQ DKEIEERSI RSTPKSIGD
 101 DLENREGSD APEGILTHGK HRLPIPEWT KEDIAAISEA AARLANPSP
 151 EDIETPSMW PYDEIRHY MOSLERKLP PPNYVQOE IDWTRHMLV
 201 DWIVQVIRH RLPLEFLA VNLIDFLSI KVSLSQVQL VGISALLINC
 251 KYEIHPSPI YNEAVVOGI FVDEILIRAE RYMLLDFD ISWPGNSFL
 301 RIRIRASVD HDIRMLAKYL QEVTLDEIF IGAHSIFLTA TAYLSMOML
 351 GHDMTQCHV YSGYTARQL KPCANIWEC LVDAPNHNA IYRKYSENRM
 401 KVSAPFAHNV YLSVI
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 196 AA.
 ID CH19_DROGR
 AC P13427;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE CHORION PROTEIN S19.
 GN CPL9 OR S19.
 OS Drosophila grimshawi (Fruit fly) (Idiomya grimshawi).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88297142.
 RA MARTINEZ-CRUZADO J.C., SWIMER C., FENERJIAN M.G., KARATOS F.C.;
 RL "Evolution of the autosomal chorion locus in Drosophila. I. General
 RL organization of the locus and sequence comparisons of genes S15 and
 RL S19 in evolutionary distant species."
 RL Genetics 119:663-677(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; X53422; CAA37506.1; -

DR PIR: S06613; S06613.
 DR FLYBASE: FBgn0012320; Dgr1Cp19.
 KW Chorion
 SQ SEQUENCE 196 AA; 19826 MW; 135544F5 CRC32;
 CH19_DMOGR Length: 196 February 14, 2000 08:02 Type: P Check: 4664 ..

1 MMTFATLAI ISACLAAGSC GGYSPIGYG GPINGLRRV SSIGQSGD
 51 AAAAASAAG GDNGPEYIIA GGAPRYGSSQ NRPILLNSG YHGGLNDWG
 101 RIAOIVGGGR SLGHLGHL GGHGCRIGG NYGGRYRPR FTVOGACATL
 151 LYPGNSYR ISSPEYSKV ILPVRAAPV AKLYIPNNY GSQVGY
 11AA_SEQUENCE 1.0 STANDARD: PRT: 213 AA.
 ID CHRR_RHOSH
 AC P0665;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL ACTIVATOR CHRR.
 GN CHRR.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 RN Rhodobacter.
 RP SEQUENCE FROM N.A.
 RX STRAIN-2.4.1;
 RX MEDLINE: 95238260.
 RA SCHIRKE B.A., DONOHUE T.J.;
 RT "ChR positively regulates transcription of the Rhodobacter
 sphaeroides cytochrome c2 gene."
 RL J. Bacteriol. 177:1929-1937(1995).
 RN [2]
 RP REVISIONS.
 RA NEWMAN J., DONOHUE T.J.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: POSITIVELY REGULATES TRANSCRIPTION OF THE CYC2 GENE
 FOR CYTOCHROME C2. MAY PLAY A ROLE IN THE INACTIVATION OF THE POE
 GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U11283; AAB17905.1;
 CC KW Transcription regulation; Activator; DNA-binding.
 FT VARIANT 38 38 C -> R (IN CHRR MUTANT; LOSS OF
 FUNCTION).
 SQ SEQUENCE 213 AA; 22865 MW; 0AAB7676 CRC32;
 CHRR_RHOSH Length: 213 February 14, 2000 08:02 Type: P Check: 8138 ..

01-NOV-1995 (Rel. 32, Last annotation update)
 DE COAT PROTEIN.
 OS Turnip crinkle virus (TCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
 OC Carnovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87283926.
 RA STOCKLEY P.G., MORRIS T.J.;
 RT "Structure and assembly of turnip crinkle virus. IV. Analysis of the
 coat protein gene and implications of the subunit primary
 structure."
 RL J. Mol. Biol. 194:265-276(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89243179.
 RA CARRINGTON J.C., HEATON L.A., ZUIDEMA D., HILLMAN B.I., MORRIS T.J.;
 RT "The genome structure of turnip crinkle virus."
 RL Virology 170:219-226(1989).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
 RX MEDLINE: 87112726.
 RA HOGUE J.M., MAEDA A., HARRISON S.C.;
 RT "Structure and assembly of turnip crinkle virus. I. X-ray
 crystallographic structure analysis at 3.2-A resolution."
 RL J. Mol. Biol. 191:625-638(1986).
 CC -1- SIMILARITY: COAT PROTEINS FROM CARNOVIRUSES (CARNV, MNSV, TCV),
 DIANTHOVIRUSES (RCMV), SOBEMOVIRUSES (MCV, SBMV), TOMBSVIRUSES
 (AMCV, CNV, CYRSV, AND TBSV), AND TNV ARE RELATED IN THEIR S
 REGION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X05193; CA28823.1;
 CC DR EMBL: M22445; AA96971.1;
 DR PIR: J01111; VCVETC.
 DR HSP: P11785; 21BY.
 DR PROSITE: P80053; ICOSAH_VIR_COAT_S; 1.
 DR PFAM: PF00729; Viral_coat; 1.
 KW Coat protein.
 FT DOMAIN 1 80 R DOMAIN, INTERACTION WITH RNA.
 FT DOMAIN 82 238 S DOMAIN, VIRION SHELL.
 FT DOMAIN 239 351 P DOMAIN, PROJECTING.
 FT CONFLICT 252 252 D -> Q (IN REF. 2).
 SQ SEQUENCE 351 AA; 38052 MW; 064C2069 CRC32;
 COAT_TCV Length: 351 February 14, 2000 08:02 Type: P Check: 1818 ..

1 MENDPRVKE ASDGAOMAIK WOKKWSITL SROQOTARAA MGKISPVQ
 51 PVQKTRLSA PVALAYREVS TOPRVSTAND GTRSGSELI TTIKNTDT
 101 PRYTTAVLNP SEPTFNQLI KEAAYEKYR FTSRFRYSP MSPSTTGKV
 151 ALAFEDDAK PPPNDLASLY NIEGCVSSVP WTGFLITVPT DSTDFVADG
 201 ISDPKLVGDP KLIMATYGOG ANDAQLGEV RVEITYOLKN RIGSTSDADI
 251 GDFAGYKQDP RLVWSKTKG TAGEMDCHF LGTGNFSITL FYKRAVSGL
 301 ENRNASDFSV LGEAAGSVQ WAGVKAERG QGVKMTTEE QPQKLOALR
 351 I
 11AA_SEQUENCE 1.0 STANDARD: PRT: 237 AA.
 ID COAT_TBSV

AC P03598: 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE COAT PROTEIN (P4 PROTEIN)
 OS Tobacco streak virus (strain WC) (TSV)
 CC Viruses: ssRNA positive-strand viruses; no DNA stage; Bromoviridae;
 CC Ilarivirus.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 84169544.
 RX CORNELISSEN B.J.C., JANSSEN H., ZUIDEMA D., BOL J.F.;
 RT "Complete nucleotide sequence of tobacco streak virus RNA 3."
 RL Nucleic Acids Res. 12:2427-2437(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X00435; CAA25133.1; .
 DR PIR: A04206; VCBWVC.
 DR SEQUENCE 237 AA; 26237 MW; 7337DEE4 CRC32;
 SQ COAT_PROSV Length: 237 February 14, 2000 08:02 Type: P Check: 3595 ..
 1 MNTLIQCPDH PSNASSSRAN NRSNNSRCP CIDEIDAMAR NCPAHTVNT
 51 VSRQRNMAA RAAAYRNANA RVPLPLPVVS VSRPOAKASL RLPNQVWYV
 101 RKASEMSAKT VDTNDAPRK TIVGCIPEIG AETKEFRLLI GFVAVSDGIF
 151 GMDVGTGDV IPPPPVGRGL GFKNTYRSR DEDLGKILN QLDRAVYWC
 201 LDERRREKRR VOLAGYWIAI SKRPAPMPPE DFLVND
 11AA_SEQUENCE 1.0 STANDARD: PRT; 359 AA.
 ID COBT_ECOLI
 AC P36562; F78075; .
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE NICOTINATE-NICOTINATE-DIMETHYLBENZIMIDAZOLE PHOSPHORIBOSYLTRANSFERASE
 CC (EC 2.4.2.21) (NN:DBI PRT) (NI-ALPHA-PHOSPHORIBOSYLTRANSFERASE).
 GN COBT.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12 / W3110;
 RX MEDLINE: 96062218.
 RX LAWRENCE J.G., ROTH J.R.;
 RT "The cobalamin (coenzyme B12) biosynthetic genes of Escherichia
 RT coli.";
 RL J. Bacteriol. 177:6371-6380(1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RX BLATTNER F.R., BLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1233-1238(1997).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12;
 RX MEDLINE: 97251358.

RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,
 RA ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,
 RA MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,
 RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
 RA SAITO N., SANPET G., SEKI Y., SIVASUNDARAM S., TAKAMI H.,
 RA TAKEDA J., TAKEKOTO K., WADA C., YAMAMOTO Y., HORUCHI T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 RN [4]
 RN SEQUENCE OF 1-188 FROM N.A.
 RP STRAIN-K12 / DHS-ALPHA;
 RC COLLINS C.M., GUTMAN D.M., ISAZA J.;
 RL Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYSES THE SYNTHESIS OF 5,6-DIMETHYLBENZIMIDAZOLE
 CC (DMB), THE LOWER LIGAND OF B12 AND THE "TRANSFER OF RIBOSE
 CC PHOSPHATE FROM NICOTINATE MONONUCLEOTIDE TO DMB.
 CC -1- CATALYTIC ACTIVITY: BETA-NICOTINATE D-RIBONUCLEOTIDE +
 CC DIMETHYLBENZIMIDAZOLE - NICOTINATE + NI-(5-PHOSPHO-ALPHA-D-
 CC RIBOSYL)-5,6-DIMETHYLBENZIMIDAZOLE.
 CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS, PART II AND III (COBII / COBIII).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U3333; AA78908.1; .
 DR EMBL: A8000291; AAC75052.1; .
 DR EMBL: D90837; CAB21728.1; .
 DR EMBL: D90838; CAB21739.1; .
 DR EMBL: L25054; AAA56876.1; .
 DR EMBL: L25055; AAA56876.1; .
 DR EMBL: L25056; AAA56876.1; .
 DR EMBL: L25057; AAA56876.1; .
 DR EMBL: L25058; AAA56876.1; .
 DR EMBL: L25059; AAA56876.1; .
 DR EMBL: L25060; AAA56876.1; .
 DR EMBL: L25061; AAA56876.1; .
 DR EMBL: L25062; AAA56876.1; .
 DR EMBL: L25063; AAA56876.1; .
 DR EMBL: L25064; AAA56876.1; .
 DR EMBL: L25065; AAA56876.1; .
 DR EMBL: L25066; AAA56876.1; .
 DR EMBL: L25067; AAA56876.1; .
 DR EMBL: L25068; AAA56876.1; .
 DR EMBL: L25069; AAA56876.1; .
 DR EMBL: L25070; AAA56876.1; .
 DR EMBL: L25071; AAA56876.1; .
 DR EMBL: L25072; AAA56876.1; .
 DR EMBL: L25073; AAA56876.1; .
 DR EMBL: L25074; AAA56876.1; .
 DR EMBL: L25075; AAA56876.1; .
 DR EMBL: L25076; AAA56876.1; .
 DR EMBL: L25077; AAA56876.1; .
 DR EMBL: L25078; AAA56876.1; .
 DR EMBL: L25079; AAA56876.1; .
 DR EMBL: L25080; AAA56876.1; .
 DR EMBL: L25081; AAA56876.1; .
 DR EMBL: L25082; AAA56876.1; .
 DR EMBL: L25083; AAA56876.1; .
 DR EMBL: L25084; AAA56876.1; .
 DR EMBL: L25085; AAA56876.1; .
 DR EMBL: L25086; AAA56876.1; .
 DR EMBL: L25087; AAA56876.1; .
 DR EMBL: L25088; AAA56876.1; .
 DR EMBL: L25089; AAA56876.1; .
 DR EMBL: L25090; AAA56876.1; .
 DR EMBL: L25091; AAA56876.1; .
 DR EMBL: L25092; AAA56876.1; .
 DR EMBL: L25093; AAA56876.1; .
 DR EMBL: L25094; AAA56876.1; .
 DR EMBL: L25095; AAA56876.1; .
 DR EMBL: L25096; AAA56876.1; .
 DR EMBL: L25097; AAA56876.1; .
 DR EMBL: L25098; AAA56876.1; .
 DR EMBL: L25099; AAA56876.1; .
 DR EMBL: L25100; AAA56876.1; .
 DR EMBL: L25101; AAA56876.1; .
 DR EMBL: L25102; AAA56876.1; .
 DR EMBL: L25103; AAA56876.1; .
 DR EMBL: L25104; AAA56876.1; .
 DR EMBL: L25105; AAA56876.1; .
 DR EMBL: L25106; AAA56876.1; .
 DR EMBL: L25107; AAA56876.1; .
 DR EMBL: L25108; AAA56876.1; .
 DR EMBL: L25109; AAA56876.1; .
 DR EMBL: L25110; AAA56876.1; .
 DR EMBL: L25111; AAA56876.1; .
 DR EMBL: L25112; AAA56876.1; .
 DR EMBL: L25113; AAA56876.1; .
 DR EMBL: L25114; AAA56876.1; .
 DR EMBL: L25115; AAA56876.1; .
 DR EMBL: L25116; AAA56876.1; .
 DR EMBL: L25117; AAA56876.1; .
 DR EMBL: L25118; AAA56876.1; .
 DR EMBL: L25119; AAA56876.1; .
 DR EMBL: L25120; AAA56876.1; .
 DR EMBL: L25121; AAA56876.1; .
 DR EMBL: L25122; AAA56876.1; .
 DR EMBL: L25123; AAA56876.1; .
 DR EMBL: L25124; AAA56876.1; .
 DR EMBL: L25125; AAA56876.1; .
 DR EMBL: L25126; AAA56876.1; .
 DR EMBL: L25127; AAA56876.1; .
 DR EMBL: L25128; AAA56876.1; .
 DR EMBL: L25129; AAA56876.1; .
 DR EMBL: L25130; AAA56876.1; .
 DR EMBL: L25131; AAA56876.1; .
 DR EMBL: L25132; AAA56876.1; .
 DR EMBL: L25133; AAA56876.1; .
 DR EMBL: L25134; AAA56876.1; .
 DR EMBL: L25135; AAA56876.1; .
 DR EMBL: L25136; AAA56876.1; .
 DR EMBL: L25137; AAA56876.1; .
 DR EMBL: L25138; AAA56876.1; .
 DR EMBL: L25139; AAA56876.1; .
 DR EMBL: L25140; AAA56876.1; .
 DR EMBL: L25141; AAA56876.1; .
 DR EMBL: L25142; AAA56876.1; .
 DR EMBL: L25143; AAA56876.1; .
 DR EMBL: L25144; AAA56876.1; .
 DR EMBL: L25145; AAA56876.1; .
 DR EMBL: L25146; AAA56876.1; .
 DR EMBL: L25147; AAA56876.1; .
 DR EMBL: L25148; AAA56876.1; .
 DR EMBL: L25149; AAA56876.1; .
 DR EMBL: L25150; AAA56876.1; .
 DR EMBL: L25151; AAA56876.1; .
 DR EMBL: L25152; AAA56876.1; .
 DR EMBL: L25153; AAA56876.1; .
 DR EMBL: L25154; AAA56876.1; .
 DR EMBL: L25155; AAA56876.1; .
 DR EMBL: L25156; AAA56876.1; .
 DR EMBL: L25157; AAA56876.1; .
 DR EMBL: L25158; AAA56876.1; .
 DR EMBL: L25159; AAA56876.1; .
 DR EMBL: L25160; AAA56876.1; .
 DR EMBL: L25161; AAA56876.1; .
 DR EMBL: L25162; AAA56876.1; .
 DR EMBL: L25163; AAA56876.1; .
 DR EMBL: L25164; AAA56876.1; .
 DR EMBL: L25165; AAA56876.1; .
 DR EMBL: L25166; AAA56876.1; .
 DR EMBL: L25167; AAA56876.1; .
 DR EMBL: L25168; AAA56876.1; .
 DR EMBL: L25169; AAA56876.1; .
 DR EMBL: L25170; AAA56876.1; .
 DR EMBL: L25171; AAA56876.1; .
 DR EMBL: L25172; AAA56876.1; .
 DR EMBL: L25173; AAA56876.1; .
 DR EMBL: L25174; AAA56876.1; .
 DR EMBL: L25175; AAA56876.1; .
 DR EMBL: L25176; AAA56876.1; .
 DR EMBL: L25177; AAA56876.1; .
 DR EMBL: L25178; AAA56876.1; .
 DR EMBL: L25179; AAA56876.1; .
 DR EMBL: L25180; AAA56876.1; .
 DR EMBL: L25181; AAA56876.1; .
 DR EMBL: L25182; AAA56876.1; .
 DR EMBL: L25183; AAA56876.1; .
 DR EMBL: L25184; AAA56876.1; .
 DR EMBL: L25185; AAA56876.1; .
 DR EMBL: L25186; AAA56876.1; .
 DR EMBL: L25187; AAA56876.1; .
 DR EMBL: L25188; AAA56876.1; .
 DR EMBL: L25189; AAA56876.1; .
 DR EMBL: L25190; AAA56876.1; .
 DR EMBL: L25191; AAA56876.1; .
 DR EMBL: L25192; AAA56876.1; .
 DR EMBL: L25193; AAA56876.1; .
 DR EMBL: L25194; AAA56876.1; .
 DR EMBL: L25195; AAA56876.1; .
 DR EMBL: L25196; AAA56876.1; .
 DR EMBL: L25197; AAA56876.1; .
 DR EMBL: L25198; AAA56876.1; .
 DR EMBL: L25199; AAA56876.1; .
 DR EMBL: L25200; AAA56876.1; .
 DR EMBL: L25201; AAA56876.1; .
 DR EMBL: L25202; AAA56876.1; .
 DR EMBL: L25203; AAA56876.1; .
 DR EMBL: L25204; AAA56876.1; .
 DR EMBL: L25205; AAA56876.1; .
 DR EMBL: L25206; AAA56876.1; .
 DR EMBL: L25207; AAA56876.1; .
 DR EMBL: L25208; AAA56876.1; .
 DR EMBL: L25209; AAA56876.1; .
 DR EMBL: L25210; AAA56876.1; .
 DR EMBL: L25211; AAA56876.1; .
 DR EMBL: L25212; AAA56876.1; .
 DR EMBL: L25213; AAA56876.1; .
 DR EMBL: L25214; AAA56876.1; .
 DR EMBL: L25215; AAA56876.1; .
 DR EMBL: L25216; AAA56876.1; .
 DR EMBL: L25217; AAA56876.1; .
 DR EMBL: L25218; AAA56876.1; .
 DR EMBL: L25219; AAA56876.1; .
 DR EMBL: L25220; AAA56876.1; .
 DR EMBL: L25221; AAA56876.1; .
 DR EMBL: L25222; AAA56876.1; .
 DR EMBL: L25223; AAA56876.1; .
 DR EMBL: L25224; AAA56876.1; .
 DR EMBL: L25225; AAA56876.1; .
 DR EMBL: L25226; AAA56876.1; .
 DR EMBL: L25227; AAA56876.1; .
 DR EMBL: L25228; AAA56876.1; .
 DR EMBL: L25229; AAA56876.1; .
 DR EMBL: L25230; AAA56876.1; .
 DR EMBL: L25231; AAA56876.1; .
 DR EMBL: L25232; AAA56876.1; .
 DR EMBL: L25233; AAA56876.1; .
 DR EMBL: L25234; AAA56876.1; .
 DR EMBL: L25235; AAA56876.1; .
 DR EMBL: L25236; AAA56876.1; .
 DR EMBL: L25237; AAA56876.1; .
 DR EMBL: L25238; AAA56876.1; .
 DR EMBL: L25239; AAA56876.1; .
 DR EMBL: L25240; AAA56876.1; .
 DR EMBL: L25241; AAA56876.1; .
 DR EMBL: L25242; AAA56876.1; .
 DR EMBL: L25243; AAA56876.1; .
 DR EMBL: L25244; AAA56876.1; .
 DR EMBL: L25245; AAA56876.1; .
 DR EMBL: L25246; AAA56876.1; .
 DR EMBL: L25247; AAA56876.1; .
 DR EMBL: L25248; AAA56876.1; .
 DR EMBL: L25249; AAA56876.1; .
 DR EMBL: L25250; AAA56876.1; .
 DR EMBL: L25251; AAA56876.1; .
 DR EMBL: L25252; AAA56876.1; .
 DR EMBL: L25253; AAA56876.1; .
 DR EMBL: L25254; AAA56876.1; .
 DR EMBL: L25255; AAA56876.1; .
 DR EMBL: L25256; AAA56876.1; .
 DR EMBL: L25257; AAA56876.1; .
 DR EMBL: L25258; AAA56876.1; .
 DR EMBL: L25259; AAA56876.1; .
 DR EMBL: L25260; AAA56876.1; .
 DR EMBL: L25261; AAA56876.1; .
 DR EMBL: L25262; AAA56876.1; .
 DR EMBL: L25263; AAA56876.1; .
 DR EMBL: L25264; AAA56876.1; .
 DR EMBL: L25265; AAA56876.1; .
 DR EMBL: L25266; AAA56876.1; .
 DR EMBL: L25267; AAA56876.1; .
 DR EMBL: L25268; AAA56876.1; .
 DR EMBL: L25269; AAA56876.1; .
 DR EMBL: L25270; AAA56876.1; .
 DR EMBL: L25271; AAA56876.1; .
 DR EMBL: L25272; AAA56876.1; .
 DR EMBL: L25273; AAA56876.1; .
 DR EMBL: L25274; AAA56876.1; .
 DR EMBL: L25275; AAA56876.1; .
 DR EMBL: L25276; AAA56876.1; .
 DR EMBL: L25277; AAA56876.1; .
 DR EMBL: L25278; AAA56876.1; .
 DR EMBL: L25279; AAA56876.1; .
 DR EMBL: L25280; AAA56876.1; .
 DR EMBL: L25281; AAA56876.1; .
 DR EMBL: L25282; AAA56876.1; .
 DR EMBL: L25283; AAA56876.1; .
 DR EMBL: L25284; AAA56876.1; .
 DR EMBL: L25285; AAA56876.1; .
 DR EMBL: L25286; AAA56876.1; .
 DR EMBL: L25287; AAA56876.1; .
 DR EMBL: L25288; AAA56876.1; .
 DR EMBL: L25289; AAA56876.1; .
 DR EMBL: L25290; AAA56876.1; .
 DR EMBL: L25291; AAA56876.1; .
 DR EMBL: L25292; AAA56876.1; .
 DR EMBL: L25293; AAA56876.1; .
 DR EMBL: L25294; AAA56876.1; .
 DR EMBL: L25295; AAA56876.1; .
 DR EMBL: L25296; AAA56876.1; .
 DR EMBL: L25297; AAA56876.1; .
 DR EMBL: L25298; AAA56876.1; .
 DR EMBL: L25299; AAA56876.1; .
 DR EMBL: L25300; AAA56876.1; .
 DR EMBL: L25301; AAA56876.1; .
 DR EMBL: L25302; AAA56876.1; .
 DR EMBL: L25303; AAA56876.1; .
 DR EMBL: L25304; AAA56876.1; .
 DR EMBL: L25305; AAA56876.1; .
 DR EMBL: L25306; AAA56876.1; .
 DR EMBL: L25307; AAA56876.1; .
 DR EMBL: L25308; AAA56876.1; .
 DR EMBL: L25309; AAA56876.1; .
 DR EMBL: L25310; AAA56876.1; .
 DR EMBL: L25311; AAA56876.1; .
 DR EMBL: L25312; AAA56876.1; .
 DR EMBL: L25313; AAA56876.1; .
 DR EMBL: L25314; AAA56876.1; .
 DR EMBL: L25315; AAA56876.1; .
 DR EMBL: L25316; AAA56876.1; .
 DR EMBL: L25317; AAA56876.1; .
 DR EMBL: L25318; AAA56876.1; .
 DR EMBL: L25319; AAA56876.1; .
 DR EMBL: L25320; AAA56876.1; .
 DR EMBL: L25321; AAA56876.1; .
 DR EMBL: L25322; AAA56876.1; .
 DR EMBL: L25323; AAA56876.1; .
 DR EMBL: L25324; AAA56876.1; .
 DR EMBL: L25325; AAA56876.1; .
 DR EMBL: L25326; AAA56876.1; .
 DR EMBL: L25327; AAA56876.1; .
 DR EMBL: L25328; AAA56876.1; .
 DR EMBL: L25329; AAA56876.1; .
 DR EMBL: L25330; AAA56876.1; .
 DR EMBL: L25331; AAA56876.1; .
 DR EMBL: L25332; AAA56876.1; .
 DR EMBL: L25333; AAA56876.1; .
 DR EMBL: L25334; AAA56876.1; .
 DR EMBL: L25335; AAA56876.1; .
 DR EMBL: L25336; AAA56876.1; .
 DR EMBL: L25337; AAA56876.1; .
 DR EMBL: L25338; AAA56876.1; .
 DR EMBL: L25339; AAA56876.1; .
 DR EMBL: L25340; AAA56876.1; .
 DR EMBL: L25341; AAA56876.1; .
 DR EMBL: L25342; AAA56876.1; .
 DR EMBL: L25343; AAA56876.1; .
 DR EMBL: L25344; AAA56876.1; .
 DR EMBL: L25345; AAA56876.1; .
 DR EMBL: L25346; AAA56876.1; .
 DR EMBL: L25347; AAA56876.1; .
 DR EMBL: L25348; AAA56876.1; .
 DR EMBL: L25349; AAA56876.1; .
 DR EMBL: L25350; AAA56876.1; .
 DR EMBL: L25351; AAA56876.1; .
 DR EMBL: L25352; AAA56876.1; .
 DR EMBL: L25353; AAA56876.1; .
 DR EMBL: L25354; AAA56876.1; .
 DR EMBL: L25355; AAA56876.1; .
 DR EMBL: L25356; AAA56876.1; .
 DR EMBL: L25357; AAA56876.1; .
 DR EMBL: L25358; AAA56876.1; .
 DR EMBL: L25359; AAA56876.1; .
 DR EMBL: L25360; AAA56876.1; .
 DR EMBL: L25361; AAA56876.1; .
 DR EMBL: L25362; AAA56876.1; .
 DR EMBL: L25363; AAA56876.1; .
 DR EMBL: L25364; AAA56876.1; .
 DR EMBL: L25365; AAA56876.1; .
 DR EMBL: L25366; AAA56876.1; .
 DR EMBL: L25367; AAA56876.1; .
 DR EMBL: L25368; AAA56876.1; .
 DR EMBL: L25369; AAA56876.1; .
 DR EMBL: L25370; AAA56876.1; .
 DR EMBL: L25371; AAA56876.1; .
 DR EMBL: L25372; AAA56876.1; .
 DR EMBL: L25373; AAA56876.1; .
 DR EMBL: L25374; AAA56876.1; .
 DR EMBL: L25375; AAA56876.1; .
 DR EMBL: L25376; AAA56876.1; .
 DR EMBL: L25377; AAA56876.1; .
 DR EMBL: L25378; AAA56876.1; .
 DR EMBL: L25379; AAA56876.1; .
 DR EMBL: L25380; AAA56876.1; .
 DR EMBL: L25381; AAA56876.1; .
 DR EMBL: L25382; AAA56876.1; .
 DR EMBL: L25383; AAA56876.1; .
 DR EMBL: L25384; AAA56876.1; .
 DR EMBL: L25385; AAA56876.1; .
 DR EMBL: L25386; AAA56876.1; .
 DR EMBL: L25387; AAA56876.1; .
 DR EMBL: L25388; AAA56876.1; .
 DR EMBL: L25389; AAA56876.1; .
 DR EMBL: L25390; AAA56876.1; .
 DR EMBL: L25391; AAA56876.1; .
 DR EMBL: L25392; AAA56876.1; .
 DR EMBL: L25393; AAA56876.1; .
 DR EMBL: L25394; AAA56876.1; .
 DR EMBL: L25395; AAA56876.1; .
 DR EMBL: L25396; AAA56876.1; .
 DR EMBL: L25397; AAA56876.1; .
 DR EMBL: L25398; AAA56876.1; .
 DR EMBL: L25399; AAA56876.1; .
 DR EMBL: L25400; AAA56876.1; .
 DR EMBL: L25401; AAA56876.1; .
 DR EMBL: L25402; AAA56876.1; .
 DR EMBL: L25403; AAA56876.1; .
 DR EMBL: L25404; AAA56876.1; .
 DR EMBL: L25405; AAA56876.1; .
 DR EMBL: L25406; AAA56876.1; .
 DR EMBL: L25407; AAA56876.1; .
 DR EMBL: L25408; AAA56876.1; .
 DR EMBL: L25409; AAA56876.1; .
 DR EMBL: L25410; AAA56876.1; .
 DR EMBL: L25411; AAA56876.1; .
 DR EMBL: L25412; AAA56876.1; .
 DR EMBL: L25413; AAA56876.1; .
 DR EMBL: L25414; AAA56876.1; .
 DR EMBL: L25415; AAA56876.1; .
 DR EMBL: L25416; AAA56876.1; .
 DR EMBL: L25417; AAA56876.1; .
 DR EMBL: L25418; AAA56876.1; .
 DR EMBL: L25419; AAA56876.1; .
 DR EMBL: L25420; AAA56876.1; .
 DR

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89106272.
 RA PHEIPS L.G., BURKE J.M., ULLRICH R.C., NOVOTNY C.P.;
 RT "Nucleotide base sequence of the mitochondrial *coII* gene of
 Schizosaccharomyces pombe."
 RT Schizosaccharomyces pombe.
 RT Curr. Genet. 14:401-403(1988).
 CC -1- FUNCTION: SUBUNIT I, II, AND III FORM THE FUNCTIONAL CORE OF
 CC THE ENZYME COMPLEX.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M36270; CAB25813.1; -
 DR PIR: S07557; S07557.
 DR HSSP: P00415; 10CC.
 DR PFAM: PF00510; COX3; 1.
 DR Oxidoreductase; Mitochondrion; Transmembrane.
 KW Oxidoreductase; Mitochondrion; Transmembrane.
 SQ SEQUENCE 268 AA; 29922 MW; 25E315EF CRC32;
 COX3_SCHCO Length: 268 February 14, 2000 08:02 Type: P Check: 7659 ..
 1 MTSIKFYQSF SAHLVQSPW PLVFSFLFN LAIGTVLNM GYSHSSTFD
 51 LGAVTVGSI LMTNDIVIE GSFLEDRTKO VQGLIGTI LFISEVAF
 101 ISVFWAFHS ALSPAVELGS TWPVGGIPL DTFSLPLNT ILLSSGAFV
 151 TYGHNAIFSG KRLLDSITGLF LTVALLAIFS YFOAFETIHA GFSMSDSVFG
 201 TYFFASTGLH GIHWMLGTLF LEVSELRQVN YQTKENHNG LETSILYWHF
 251 VDLVWLEFL VYFWGGA
 !!AA-SEQUENCE 1.0
 ID COX3_SCHCO STANDARD: PRT: 164 AA.
 AC P79010;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE IV PRECURSOR (EC 1.9.3.1).
 GN COX4.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KAWAMUKAI M.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
 CC -1- SUBUNIT: COMPOSED OF AT LEAST 11 SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB000399; BAA19097.1; -

DR PROSITE: PS00848; COX5B; 1.
 DR PFAM: PF01215; COX5B; 1.
 KW Oxidoreductase; Mitochondrion; Transist peptide; Zinc.
 FT TRANSIT 1 33
 FT CHAIN 1 33
 FT METAL 34 164
 FT METAL 120 120
 FT METAL 120 120
 FT METAL 143 143
 FT METAL 146 146
 FT METAL 146 146
 SQ SEQUENCE 164 AA; 18161 MW; DA19B411 CRC32;
 COX4_SCHCO Length: 164 February 14, 2000 08:02 Type: P Check: 7810 ..
 1 MEMNSMLRVS RORAAVSTV SLVRFVSAS IRREQNVVK AAOELANAK
 51 EPSDLIGPGG RDEVEPTDLE QATGDERVEL LSELGSDAF DKPPLDASK
 101 GLTIDPIWVT SLDPYRHICG TGSPGSHNL IMWTVYKDKL RCPGCGSVY
 151 KLFKMGDPNA EHSK

!!AA-SEQUENCE 1.0
 ID CPCL_CANGP STANDARD: PRT: 114 AA.
 AC P81580;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CUTICLE PROTEIN CP1158 (CPCL1158).
 OS Cancer pagurus (Rock crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura; Cancridae;
 OC Cancridae; Cancridae; Cancer.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-CUTICLE;
 RA ANDERSEN S.O.;
 RT "Exoskeletal proteins from the crab, Cancer pagurus.";
 RT Comp. Biochem. Physiol. 0:0-0(1999).
 CC -1- TISSUE SPECIFICITY: CALCIFIED SHELL.
 KW Structural protein; Cuticle; Repeat.
 FT MOD_RES 1 1
 FT REPEAT 1 17
 FT REPEAT 26 43
 FT REPEAT 70 87
 FT REPEAT 95 112
 SQ SEQUENCE 114 AA; 11601 MW; FC0D861C CRC32;
 CPCL_CANGP Length: 114 February 14, 2000 08:02 Type: P Check: 1605 ..
 1 QVYSGIVSP DGNICPTFD FAHSIVLNGP SGIVTSQKN LQITAGQASL
 51 QAAAPAPLP VSHVVAQGS VVGPSGIYSP SGNVQFSHER ADNVVLYVPS
 101 GIVTKDGNL QLRA
 !!AA-SEQUENCE 1.0
 ID CPTR_PEA STANDARD: PRT: 402 AA.
 AC P21727;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TRIOSE PHOSPHATE/PHOSPHATE TRANSLATOR, CHLOROPLAST PRECURSOR (CPPT)
 DE (P36) (E00)
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 73-98; 249-260; 269-289 & 324-350.
 RX STRAIN: CV. PROGRESS NO. 9; TISSUE-SEEDLING;
 RA MEDLINE: 91035597.
 RA SCHNELL D.J., BLOBEL G., PAIN D.;
 RT "The chloroplast import receptor is an integral membrane protein of

RT	chloroplast envelope contact sites."
U Cell Biol.	111:1825-1838(1990).
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 73-82.
RA	TISSE-LEAF; FISCHER K., WACHTER E., LINK T.A., FLUEGE U.-I.;
RT	Molecular cloning and structural analysis of the phosphate
RT	translocator from pea chloroplasts and its comparison to the spinach
RT	phosphate translocator."?
PLanta	183:451-461(1991).
[3]	
RN	SEQUENCE FROM N.A.
RP	WEBER A.;
RL	Submitted (SEP-1992) to the EMBL/genbank/DBJ databases.
CC	-1- FUNCTION: MEDIATES THE EXPORT OF FIXED CARBONS FROM THE
CC	CHLOROPLASTS INTO THE CYTOSOL IN THE FORM OF TRIOSE PHOSPHATES.
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LOCATED IN
CC	ZONES OF CONTACT BETWEEN THE INNER AND OUTER MEMBRANE OF THE
CC	CHLOROPLAST.
CC	-1- SIMILARITY: STRONG, TO OTHER PLANTS CTPP. ALSO SIMILAR TO
CC	YEAST PROTEIN SLV4.
CC	-1- CAUTION: WAS ORIGINALLY THOUGHT TO FUNCTION AS A CHLOROPLAST
CC	PROTEIN IMPORT RECEPTOR.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X54639; CAN8451.1; ..
DR	PIR; S23774; S23774.
DR	PIR; S23774; S23774.
KW	Transmembrane; Chloroplast; Transit peptide; Transport.
FT	TRANSIT 1 72 CHLOROPLAST.
FT	CHAIN 73 402 TRIOSE PHOSPHATE/PHOSPHATE TRANSLOCATOR.
FT	DOMAIN 73 96 INTERMEMBRANE SPACE (POTENTIAL).
FT	TRANSMEM 97 117 POTENTIAL.
FT	DOMAIN 118 129 LUMENAL (POTENTIAL).
FT	TRANSMEM 130 150 INTERMEMBRANE SPACE (POTENTIAL).
FT	DOMAIN 151 207 POTENTIAL.
FT	TRANSMEM 208 228 LUMENAL (POTENTIAL).
FT	DOMAIN 229 272 POTENTIAL.
FT	TRANSMEM 273 292 INTERMEMBRANE SPACE (POTENTIAL).
FT	DOMAIN 293 370 POTENTIAL.
FT	TRANSMEM 371 391 LUMENAL (POTENTIAL).
FT	DOMAIN 392 402 LUMENAL (POTENTIAL).
FT	VARIANT 227 227 I->L
SQ	SEQUENCE 402 AA; 43649 MW; ACCIEF795 CRC32;
CPR_PEA	Length: 402 February 14, 2000 08:02 Type: P Check: 5889 ..
1	MESRVLSRAT TLSSUPTLNK LHRLLPLANAS LPVSXSGVY SDGNNLYMGR
51	QLARELCSPV LKKGASLRP CPATAGNDS AGEKVAPYG FFSRYPALIT
101	GFEFFTYFL NVIFNLNRK IYNFPYPYE VSYHLAVG YCLVSWTVG
151	LPRAPIDGN LKLIPAVY CHALGHVTSN VSFAVAAYSF THIVKALEPF
201	FNAASQFIL GGSIPITWL SLAPVIGVS MASLTSLFN WLGISAMIS
251	NISTYSISY SKRAMTMDMS TNIVAYISI ALIVCIPPAL IIEGPTLLKT
301	GFNDIAKVG LYKFVSDLM VGMEHYLNQ VATNTLERVA PLTHAVGVNL
351	KRVYVIGST IIFGNKRSTQ TGIGTGIAA GVAYSIRKA QEEEKROAK
401	AA

PL J. Biol. Chem. 268:678-685(1993).
 CC -1- FUNCTION: CTA is REQUIRED FOR CYTOCHROME A3 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M94110; AAA22362.1; -
 DR PIR: A45335; A45335.
 KW Sporulation; Membrane.
 SQ SEQUENCE 297 AA; 32326 MW; BA198CC1 CRC32;

CTAA_BACFI Length: 297 February 14, 2000 08:02 Type: P Check: 716

1 MKRRLKITSV ITSIGVIAL LOGALVTKTG SGEGCATWP LCFGEVTPN
 51 PAETIIYS HRIVSLAGA MIITAIAM KQKHMEKAL ALFAAVILI
 101 ISQGLLAGA AVFGQSKAIL ALHFGISAMS LAHVLLITLI AFEDGREHTM
 151 APKVSRGKY YVEFVITTCY AVTSGAYK HSEATACAG PFLNGOITP
 201 GLYGVGMY FHRYVGTILL LFLILMIT LSRHYRYVL TWTAVLSFL
 251 VVGOFISIS IYTONALSV GLIHALLISI LFSALSYMT IITPSH
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 306 AA.
 ID CTA_BACFI
 AC P12946
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE CRYOCHROME A3 CONTROLLING PROTEIN.
 OS CTA.
 CC Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89359135.
 RA MOELLER J.P., TABER H.W.;
 RT "Isolation and sequence of cta, a gene required for cytochrome a3
 RT biosynthesis and sporulation in Bacillus subtilis."
 RL J. Bacteriol. 171:4967-4978(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168;
 RA BERTERO M., PRESBAN E., GLASER P., RICHOT A., DANCHIN A.;
 RT Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CTA is REQUIRED FOR CYTOCHROME A3 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23915; AAA50254.1; -
 DR EMBL: 298682; CAB1340.1; -
 DR EMBL: 298682; CAB1340.1; -
 DR PIR: A33960; A33960.
 DR SUBLIST: Bg10213; CTA.
 KW Sporulation; Membrane.

SQ SEQUENCE 306 AA; 34085 MW; 366EB369 CRC32;
 CTAA_BACSV Length: 306 February 14, 2000 08:02 Type: P Check: 1522

1 MKRALKAV LTFVMDIV ICGALVTKTG SGQCGGQWP LCHGRFPPEL
 51 NPASIEWSH RFASGISIIL VSLAFWSMR KIPIPRETT FLAISIIFL
 101 FLQALIGALA VFGSNALIM ALHFGISLIS FASVILLITLI IFADKSVRT
 151 LVKPLQIGRK MEFMIGILI YSYIVYGA YRHTESSILA CPNPPLCSPL
 201 NNGLPLOFHE WQNGHRAA LLFWIIVA AVHAITSYD QKQIFGWNIS
 251 CLIFITLQAL SGIMITYSEL ALGFALHMF FICLFGVLC YFLLIARFR
 301 YESROS

!!AA_SEQUENCE 1.0 STANDARD; PRT; 127 AA.
 ID CU04_BACR
 AC P80675;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE CURTICLE PROTEIN 4 (BC-NC4).
 OS Blaberus craniifer.
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blaberoidea; Blaberidae; Blaberus.
 RN [1]
 RP SEQUENCE
 RC TISSUE-POLYRIC CAECA;
 RX MEDLINE: 97218697.
 RA JENSEN U.G., ROTHMANN A., SKOU L., ANDERSEN S.O., ROEPSTORFF P.,
 RA HOLZBUP P.;
 RT "Cuticular proteins from the giant cockroach, Blaberus craniifer."
 RL Insect Biochem. Mol. Biol. 27:109-120(1997).
 KW Structural protein; Cuticle; Repeat.
 FT MOD_RES 1 1
 FT REPEAT 31 39
 FT REPEAT 84 92
 FT SEQUENCE 127 AA; 13383 MW; 26A82D7C CRC32;
 CU04_BACR Length: 127 February 14, 2000 08:02 Type: P Check: 1926

1 QAVLPSIYS ILSKSVRYQ ELEPYEGAV PADTKAEIK EKEHEVITHG
 51 LPVPSAVQ PIVSVVNP GVIPAVPYD APIDATKKA EISDNVAVY
 101 HAISPLSVV SHPVSPVH VQVYVY
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 92 AA.
 ID CY2_RHOE
 AC P00098;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE CYTOCHROME C2.
 OS Rhodocyclis tenuis (Rhodospirillum tenue).
 CC Bacteria; Proteobacteria; beta subdivision; Rhodocyclis group;
 CC Rhodocyclis.
 RN [1]
 RP SEQUENCE.
 RC STRAIN-3761;
 RX MEDLINE: 79199668.
 RA AMBLER R.P., MEYER T.E., KAMEN M.D.;
 RT "Anomolies in amino acid sequences of small cytochromes c and
 RT cytochromes c' from two species of purple photosynthetic bacteria."
 RL Nature 278:661-662(1979).
 CC -1- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR
 CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR
 CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
 CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN

AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.
 -1- SIMILARITY: THIS SEQUENCE IS MORE CLOSELY RELATED TO THE SEQUENCES OF CYTOCHROME C551 FROM PSEUDOMONAS AND AZOBACTER THAN TO THE SEQUENCES OF CYTOCHROME C2 FROM OTHER SPECIES OF RHODOSPIRILLUM.
 PIR: A00090; CCOR2T.
 DR HSP: P00099; 2PAC.
 DR PROSITE: P500190; CYTOCHROME_C; 1.
 DR PFAM: PF00034; cytochrome_c; 1.
 DR Electron transport; Photosynthesis; Heme.
 KW BINDING 12 12 HEME (COVALENT).
 FT BINDING 15 15 HEME (COVALENT). LIGAND.
 FT METAL 16 16 IRON (HEME AXIAL LIGAND).
 FT METAL 66 66 IRON (HEME AXIAL LIGAND).
 FT SEQUENCE 92 AA; 9781 MW; 573C3CF0 CRC32;
 CY2_RHOTE Length: 92 February 14, 2000 08:02 Type: P Check: 2620 ..

1 ADESALAQTK GCLACHNPEK KVGQPAYGVV AKRYAGAGA EAKLVAKVNA
 51 GGQGVMAKOL GAEIPMPANN VTKREATLV KWTLSKQID YK

11AA SEQUENCE 1.0 STANDARD: PRT: 299 AA.
 ID CYNR_ECOLI
 AC P27111: 1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CYN OPERON TRANSCRIPTIONAL ACTIVATOR.
 GN CYN.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 92276346.
 RA SUNG Y.-C., FUCHS J.A.;
 RT "The Escherichia coli K-12 cyn operon is positively regulated by a member of the lysR family";
 RT J. Bacteriol. 174:3645-3650(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERRA N.T., BURLAND V., RILEY M., COLLADO-VIDES J., GLASNER F.D., ROBE C.K., MAYHEW G.F., GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J., MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RT Science 277:1453-1474(1997).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC DUNCAN M., ALLEN E., ARAUJO R., APARICIO A.M., CHUNG E., DAVIS K., FERREPEITL N., HYMAN R., KALMAN S., KOMP C., KURI O., LEW H., LIND D., NAMATH A., OEFNER P., ROBERTS D., SCHRAMM S., DAVIS R.W.;
 RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: POSITIVE REGULATORY PROTEIN FOR THE CYN OPERON.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- STRAINTY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M93053; AA23628.1;
 DR EMBL: AE000141; AAC73441.1;
 DR EMBL: U73857; AAB18062.1; ALT_INIT.
 DR PIR: A41900; A41900.

ECOGENE: EG11421; CYNR.
 DR PROSITE: P500044; HTL_LYSR_FAMILY; 1.
 DR PFAM: PF00126; HTL_1; 1.
 DR Transcription regulation; DNA-binding; Activator.
 KW DNABIND 18 37 H-T-H MOTIF (BY SIMILARITY).
 FT SEQUENCE 299 AA; 32961 MW; 0A8298D6 CRC32;
 CYNR_ECOLI Length: 299 February 14, 2000 08:02 Type: P Check: 4304 ..

1 MSTRINFL AVAHEGSETR AASALHVSOP ALSQIROLE ESLGVPLEDR
 51 SGRTIRLDA GEVWQYASR ALQELGAGKR AIHDVADLR GSLRIAVPT
 101 FTSYFGLPM ADFAPRPSI TLQLEMSQE KIEDMLCRDE LDVGIAFAPV
 151 HSPLEALPL LTESLALVA QHPLPLAVHQ VALSRHDEK LVLSAEFAT
 201 REQIDHYCEK AGLHPQVTE ANSISAVLEL IRRSTSLTL PAIAFQHDG
 251 LKAISLAPPL LERTAVLILR KNSWQTAANK AFLHMAIDKC AVAGNSES

11AA SEQUENCE 1.0 STANDARD: PRT: 291 AA.
 ID DACC_STRSK
 AC P39042;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DT D-ALANYL-D-ALANINE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.16.4) (DD-PEPTIDASE) (DD-CARBOXYPEPTIDASE).
 DE Streptomyces sp. (strain K15).
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Bacteri; Firmicutes; Streptomycetaceae; Streptomycetes.
 OC Actinomycetales; Streptomycetaceae; Streptomycetes.
 [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 92028812.
 RA PALOMQUE-MESSIA P., ENGELBERT S., LEYH-BOUILLE M., NGUYEN-DISTECHE M., DIEZ C., HOUBA S., DIDBERG O., VAN BREDEN J., GHUYSEN J.-M.;
 RT "Amino acid sequence of the penicillin-binding protein/penicillin-binding proteins of class A";
 RT Biochem. J. 279:223-230(1991).
 RL [2]
 RP SEQUENCE OF 30-85, AND ACTIVE SITE.
 RC MEDLINE: 89350853.
 RA LEYH-BOUILLE M., VAN BREDEN J., RENIER-PHILLOT S., JORIS B., NGUYEN-DISTECHE M., GHUYSEN J.-M.;
 RT "The streptomycetes K15 DD-peptidase/penicillin-binding protein. Active site and sequence of the N-terminal region";
 RT Biochem. J. 260:601-604(1989).
 RL [3]
 CC -1- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE CELL WALL PRECURSORS.
 CC -1- CATALYTIC ACTIVITY: D-ALANYL-D-ALANINE + H₂O = 2 D-ALANINE.
 CC -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11. ALSO KNOWN AS THE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X59965; CAA42591.1;
 DR PIR: S04638; S04638.
 DR PIR: S17674; S17674.
 DR PFAM: PF00768; Peptidase_S11; 1.
 DR Hydrolyase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall; signal.
 KW signal.
 FT CHAIN 1 29
 FT SIGNAL 30 291 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.

Mon Feb 14 08:07:43 2000

sp.cat

Page 46

ACT SITE 64 64 ACYLATED BY PENICILLIN.
FT CONFLICT 70 70 T -> Q (IN REF. 2).
FT CONFLICT 83 83 D -> S (IN REF. 2).
SQ SEQUENCE 291 AA; 30257 MW; E2D9DFE9 CRC32;

DACC_STRK Length: 291 February 14, 2000 08:02 Type: P Check: 4817

1 MRLRAATV ITTGALLAG TLGATPATAV TKPTIAVVG YAMNNGTGT
51 LYTAADTRR STGSTKIMT AKVYLAOSNL NLPKVTIK AYSDYVANK
101 PSQHLIVD KTVROLLYG LMPSCGDA YALADKSGG SOAARVASF
151 IGRANTATN LGLNTHFDS FDGIGNGANY STPRHLTKIA SSAMKSTFR
201 TVVTKAYTA KTVTKGSIR TMDTKWNTNG LLSYSGAIG VKTSGPEAK
251 YCLVPAATRG GKTVIGTVLA STSIPARESD ATKIMYGA L

!!AA_SEQUENCE 1.0
ID DBPS_YEAST STANDARD; PRT: 482 AA.
AC P20449;

DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE ATP-DEPENDENT RNA HELICASE DBPS (HELICASE CA5/6).
GN DBP5 OR YOR046C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
RN [1]
RP SEQUENCE FROM N.A.
RA CHANG T.-H.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RA LANDT O., HISEL R., UNSELD M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RA BOHN C., BOLOTIN-FUKUHARA M., DAIGNAN-FORMIER B., DANG D. V.,
VALENS M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 237-425 FROM N.A.
RX MEDLINE; 90160368.
RA CHANG T.-H., ARENAS J., ABELSON J.;
RT "Identification of five putative yeast RNA helicase genes.";
CC Proc. Natl. Acad. Sci. U.S.A. 87:1571-1575(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL; U28135; AAB01679.1;
DR EMBL; 274954; CAAG9237.1;
DR EMBL; 274955; CAAG9239.1;
DR PIR; D34848; D34848.
DR SGD; L0003292; DBP5.
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
DR PFAM; PF00270; DEAD.1.
DR PFAM; PF00271; helicase_C; 1.
KW NP_BIND 138 145 ATP (POTENTIAL).
FT SITE 239 242 DEAD BOX.
SQ SEQUENCE 482 AA; 53874 MW; C70EB665 CRC32;

DBPS_YEAST Length: 482 February 14, 2000 08:02 Type: P Check: 2958

1 MSDTRKPAD LLAALKIDNE KEDTSEVSTR ETVKSPQEPKT ADSIKPAEKL
51 VPKVEERKTK QEDSNLISE YEVKVLADI QADPNSPILS AKSDEGLA
101 PELKGIYAM KFOKPSKIOE RALPLLHNP PNMMAOSOS GTGTAAFSL
151 TMLTRVPED ASPOAICLAP SRELAROTLE VQEGKFTK ITSOLIYPS
201 FEKNKOINQ VIVGTGTVL DLMRRKLMQ OKIKFVLDE ADNMDOOGL
251 GDOCIKRAF LPRDTQVLF SATFADAVNG YAKKIYVANN TLELOINEVN
301 VDAIKOLYMD CKNEADKFDV LTELXGLMTI GSSITFAVK KLANVYKGL
351 KSEGEVSL HGDQTOERD RLIDDFRGR SKVLITTVL ARGIDIPYVS
401 MYVNYDLPTL ANQADPATY IHRIGTRGF GRKVALSEV HDKNSENILS
451 AIOKYGDI MTRVPTDMD EYKIVKYL KD

!!AA_SEQUENCE 1.0
ID DEP_STRMU STANDARD; PRT: 145 AA.
AC 054433;

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG (FRAGMENT).
GN DEP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96312359.
RA GUTIERREZ J. A., CROWLEY P. J., BROWN D. P., HILLMAN J. D.,
YOUNGMAN P., BLEIWEIS A. S.;
RT "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements.";
RL J. Bacteriol. 178:4166-4175(1996).
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL; U48885; AAC44502.1;
DR FLAVOPROTEIN.1
FT NON_TER 145
SQ SEQUENCE 145 AA; 15819 MW; 24E84944 CRC32;

DPP_STRK Length: 145 February 14, 2000 08:02 Type: P Check: 9135

1 MTKLILAVS GSIAAYKAD ISHOLTKLYG HNVNFMNAA KQFIPPLTLO
51 VLSKNPYSN VKKEDPOVI NHIALAKOAO LFLPPASAN TLALAHGHA
101 DNITSVALA LPLEVKEFA PAMNTKMYEN PIQNSNTLL KKEFG

!!AA_SEQUENCE 1.0
ID DHAT_CITER STANDARD; PRT: 387 AA.
AC P45513;
DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE 1,3-PROPANEDIOL DEHYDROGENASE (EC 1.1.1.202) (3-HYDROXYPROPYIONALDEHYDE
 DE REDUCTASE) (1,3-PROPANEDIOL OXIDOREDUCTASE).
 GN DHAT.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Citrobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 30040;
 RX MEDLINE; 95238288.
 RA DANIEL R., BOENICK R., GOTTSCHEK G.;
 RT Purification of 1,3-propanediol dehydrogenase from Citrobacter
 RT freundii and cloning, sequencing, and overexpression of the
 RT corresponding gene in Escherichia coli.";
 RL J. Bacteriol. 177:2151-2156(1995).
 CC -1- FUNCTION: MOST ACTIVE WITH SUBSTRATES CONTAINING TWO PRIMARY
 CC ALCOHOL GROUPS SEPARATED BY ONE OR TWO CARBON ATOMS. IN THE
 CC PHYSIOLOGICAL DIRECTION, 3-HYDROXYPROPYIONALDEHYDE IS THE PREFERRED
 CC SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: PROPANE-1,3-DIOL + NAD(+) = 3-HYDROXYPROPANAL
 CC + NADH.
 CC -1- COFACTOR: IRON.
 CC -1- SUBUNIT: HOMOCYMER.
 CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 009771; AAB4848.1; -
 CC DR PROSITE; P500913; ADH_IRON_1; 1.
 CC DR PROSITE; P500913; ADH_IRON_2; 1.
 CC DR PFAM; PF00465; Fe-ADH; 1.
 CC DR Oxidoreductase; NAD; Iron.
 CC SQ SEQUENCE 387 AA; 41481 MW; BAC8846 CRC32;
 DHAT_CITFR Length: 387 February 14, 2000 08:02 Type: P Check: 101 ..
 1 MSYRMEDYLY PNVNFEFPA ISVGERCKL LGKKALLVY DGLRAIKDG
 51 AVDKLTILHR EAGIDVVFV GVEPNPKDTN VRDGLVFRK EECDIIVTVG
 101 GGSFHDGCKG IGIATHEGD LYSVAGIETL TNPPLPIYAV NTAAGTASEV
 151 TRHCVLNTK TKRFEIVSW RNLPSVIND PLMLGKPA LTAATGMAL
 201 THAVEYISK DANPYTDAA IQAIRLIARN LQAVVALGNS LQARENMAVA
 251 SLGAGAFNN ANLGIVHAMA HOLGLYDMP HGVANAVILP HVARYNLIAN
 301 PEKADIAEL MGENITGLST MDAELAIAH IARLSADIGI PQLRLDGVK
 351 EADFPYAEW ALKGNAFSN PRKNGEKEIA EIFRQAF
 11AA_SEQUENCE 1.0 STANDARD; PRT; 387 AA.
 ID DHAT_KLEPN
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE 1,3-PROPANEDIOL DEHYDROGENASE (EC 1.1.1.202) (3-HYDROXYPROPYIONALDEHYDE
 DE REDUCTASE) (1,3-PROPANEDIOL OXIDOREDUCTASE).
 GN DHAT.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25955;
 RA SKRAJL F.A., WILLARD B.L., CAMERON D.C.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: PROPANE-1,3-DIOL + NAD(+) = 3-HYDROXYPROPANAL
 CC + NADH.
 CC -1- COFACTOR: IRON (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOCYMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 030903; AAA74260.1; -
 CC DR PROSITE; P500913; ADH_IRON_1; 1.
 CC DR PROSITE; P500913; ADH_IRON_2; 1.
 CC DR PFAM; PF00465; Fe-ADH; 1.
 CC DR Oxidoreductase; NAD; Iron.
 CC KM SEQUENCE 387 AA; 41465 MW; 04B25C1 CRC32;
 DHAT_KLEPN Length: 387 February 14, 2000 08:02 Type: P Check: 846 ..
 1 MSYRMEDYLY PNVNFEFPA ISVGERCKL LGKKALLVY DGLRAIKDG
 51 AVDKLTILHR EAGIDVVFV GVEPNPKDTN VRDGLVFRK EECDIIVTVG
 101 GGSFHDGCKG IGIATHEGD LYSVAGIETL TNPPLPIYAV NTAAGTASEV
 151 TRHCVLNTK TKRFEIVSW RNLPSVIND PLMLGKPA LTAATGMAL
 201 THAVEYISK DANPYTDAA IQAIRLIARN LQAVVALGNS LQARENMAVA
 251 SLGAGAFNN ANLGIVHAMA HOLGLYDMP HGVANAVILP HVARYNLIAN
 301 PEKADIAEL MGENITGLST LQAEKAIAA IRLSMDIGI PQLRLDGVK
 351 EADFPYAEW ALKGNAFSN PRKNGEKEIA AIFRQAF
 11AA_SEQUENCE 1.0 STANDARD; PRT; 310 AA.
 ID DHAT_HUMAN
 AC P37058; HUMAN
 DT 01-NOV-1994 (Rel. 29, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-DEC-1998 (Rel. 37, last annotation update)
 DE ESTRADIOL 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HYDROXYSTEROID DEHYDROGENASE).
 GN HSD17B3 OR EDH1/B3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CC [1]
 CC SEQUENCE FROM N.A., AND VARIANTS MPH Q-80; V-203; K-232 AND V-235.
 CC TISSUE-TESTIS;
 RC MEDLINE; 94355972.
 RA GEISLER W.M., DAVIS D.L., WU L., BRADSHAW K.D., PAEL S.,
 RA MENDONCA B.B., ELLISTON K.O., WILSON J.D., RUSSELL D.W.,
 RA ANDERSSON S.;
 RT "Male pseudohemaphroditism caused by mutations of testicular 17
 RT beta-hydroxysteroid dehydrogenase 3.";
 RL Nat. Genet. 7:34-39(1994).
 CC [2]
 CC VARIANT MPH TRP-80.
 RP MEDLINE; 98429412.
 RA BILBAO J.R., LORIDAN L., AUDI L., GONZALO E., CASTANO L.;
 RA "A novel missense (R80W) mutation in 17-beta-hydroxysteroid
 RT dehydrogenase type 3 gene associated with male

RT Pseudohermaphroditism.
 RL Eur. J. Endocrinol. 139:330-333(1998).
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH.
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- DISEASE: DEFECTS IN EDH17B GIVE RISE TO MALE PSEUDHERMAPHRODISM (MPH). THESE 46 XY INDIVIDUALS HAVE UNMATURED FEMALE EXTERNAL GENITALIA AT BIRTH, BUT FAIL TO MENSTRUATE AT THE TIME OF EXPECTED PUBERTY AND INSTEAD VIRILISE AS EVIDENCED BY GROWTH OF THE PHALLUS. BREAST DEVELOPMENT MAY OR MAY NOT TAKE PLACE.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U05659; AAC50066.1;
 DR MIM: 264300;
 DR PROSITE: PS00061; ADH_SHORT: 1.
 DR PFAM: PF00106; adh_short: 1.
 KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family;
 KM Pseudohermaphroditism; Disease mutation.
 FT NP_BIND 48 NADP (BY SIMILARITY).
 FT ACT_SITE 198 198 BY SIMILARITY.
 FT VARIANT 80 80 R -> Q (IN MPH; GAZA).
 FT VARIANT 80 80 R -> W (IN MPH).
 FT VARIANT 203 203 A -> V (IN MPH).
 FT VARIANT 232 232 S -> L (IN MPH).
 FT VARIANT 235 235 M -> V (IN MPH).
 FT SEQUENCE 310 AA; 34515 MW; 652F1024 CRC32;
 DHB3_HUMAN Length: 310 February 14, 2000 08:02 Type: P Check: 3434
 1 MGDVLEQFEI LTGLVCIAC LACVRFSRC VLNWVYLP KSLRSMGV
 51 AVITGAGDGI GKAYSFEIAR RGLNVVLSR TLEKLEAIAT EIERITGRSV
 101 KIIOADFTKD DIVEHIREKL AGLEIGILVN NVGMPLNLP SHLNAPDEI
 151 OSILHCNITS VKMTOLILK HMESSRQGLI LNISSGIALF PWLYIMYISA
 201 SKAFVCAFSK ALOEYKAKE VLIIVLTPEYA VSTAMTKYLN TNYITTADE
 251 FVKESLINVT IGGETGCCIA HEILNGFSL IPANAFYSGA FORLLITHTY
 301 AYLRNLTQVR
 11AA_SEQUENCE 1.0
 ID DHB3_MOUSE STANDARD; PRT; 305 AA.
 AC P70385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE (ESTRADIOL 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HYDROXYSTEROID DEHYDROGENASE).
 GN HSD17B3 OR EDH17B3.
 OS Mus musculus (mouse)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 CC SEQUENCE FROM N.A.

RC TISSUE-TESTIS;
 RA MEDLINE; 97325842.
 RX SHA J.A., DUDLEY K., RAJAPAKSHA W.R.A., O'SHAUGHNESSY P.J.;
 RT "Sequence of mouse 17beta-hydroxysteroid dehydrogenase type 3 cDNA and tissue distribution of the type 1 and type 3 isoform mRNAs."
 RL J. Steroid Biochem. Mol. Biol. 60:19-24(1997).
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U66827; AAB06793.1;
 DR MGD: MGI:107177; HSD17B3.
 DR PROSITE: PS00061; ADH_SHORT: 1.
 DR PFAM: PF00106; adh_short: 1.
 KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 KM NP_BIND 44 73 NADP (BY SIMILARITY).
 FT ACT_SITE 194 194 BY SIMILARITY.
 FT SEQUENCE 305 AA; 34271 MW; 10CB2855 CRC32;
 DHB3_MOUSE Length: 305 February 14, 2000 08:02 Type: P Check: 9317
 1 MERFLIAGL FVGLVCIYKC MRFSEHLFLR FCKALPSSFL RSMGQNAVIT
 51 GAGDGIGNAY SFELARHGLN VYLISRTLEK LQITAEIER TTGSGYKIYO
 101 ADFTREDIYD HIKHEHLEGL NGLVNVNNGM LPSTFPBSHFL SSGESQNI
 151 HCNITSYVKM TQVLKHMES RRGGLINIS SGALBRWPL YSLYSASNAF
 201 YTFESKALSY EYRDKGIID VLTPISTP MTKYLNKMT KTADEFVEX
 251 LKYVIGAES CGCLAHETIA IILNRPISRI FYSTAQRELT LTRYSDYLKR
 301 NISNR
 11AA_SEQUENCE 1.0
 ID DHB3_RAT STANDARD; PRT; 306 AA.
 AC O54939;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE (ESTRADIOL 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HYDROXYSTEROID DEHYDROGENASE).
 GN HSD17B3 OR EDH17B3.
 OS Rattus norvegicus (Rat)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 CC SEQUENCE FROM N.A.
 RA KHANUM A., TSAI-MORRIS C.-H., DUFAU M.L.;
 RT Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute has no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF035156; AB99739.1; FALSE-NEG.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 KM N3BIND 44 NADP (BY SIMILARITY).
 FT ACT_SITE 194 194 BY SIMILARITY.
 FT SEQUENCE 306 AA: 34223 MW: 38803 B: BFB8CB9 CRC32:
 DH3_RAT Length: 306 February 14, 2000 08:02 Type: P Check: 3366

1 ME0FLSYGL LVCLVCLVVC VRSRYLFLS FCKALPSEFL RSMGOMAVIT
 51 GAGDGLKAY SPELARHGLN VVLISRTLEK LQVISELIER TTGSRKVVQ
 101 ADFTREDITD HIEBOLKLE IGVLVNVMG IPNLPSHFL STGESQSVI
 151 HCNITSVYKM TOLVVKHMS RRGGLINIS SGVGRPMPL YSLYSKRAF
 201 VCFESKALNV EYRDKGIITQ VLPYVSVP MTKYINTSRV TKTADEFKE
 251 SKTYVTIGAE TCCCLAHLEL AILNLPSR IFYSSTORF LMKOFSDYK
 301 SNISNR
 11AA_SEQUENCE 1.0 STANDARD; PRT; 312 AA.
 ID DH3_MOUSE
 AC 070503;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PUTATIVE STEROID DEHYDROGENASE KIK-1 (EC 1.1.1.-).
 GN KIK1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 RN SEQUENCE FROM N.A.
 RC SPRAIN-BALB/C; TISSUE-LIVER;
 RA GAMBORTO A.; PAGLIANO O.; ROBBINS P.; DELBO A.;
 RA Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 RL -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 -1 FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute has no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF064635; AAC1685.1;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 KM NP_BIND 50 NADP (BY SIMILARITY).
 FT ACT_SITE 202 202 BY SIMILARITY.
 FT SEQUENCE 312 AA: 34742 MW: 1210554 CRC32:
 DH3_MOUSE Length: 312 February 14, 2000 08:02 Type: P Check: 4788

1 MECAPPAAGF IYWGASTIA YLALRASYSL FRAVWCYCVG NHALVGRIG
 51 EMAYVTGTD GIGKAYAEEL AKRGKIVLI SRSDKLNNOV SNRIKEKRV

101 ERTIAVDFS LDDIYDKIKT GIGLEIGVL YNNGMSYEV PEYFLPIPLD
 151 DNTIKLINI NVLSVCVTR LVLPWERS KYVLINISSA SGMLPVLLT
 201 IYATKAFVD FFSQCLHEEY KSKGIFVOS MPLVATKLA KIQPTLDKP
 251 SATFVKSAI KTVGLQTRTT GYVISHLMS INSIMPRMY FKIMGFESK
 301 LNRRLKRR KN

11AA_SEQUENCE 1.0 STANDARD; PRT; 312 AA.
 ID DH3_MOUSE
 AC 075828;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE STEROID DEHYDROGENASE SPW2 (EC 1.1.1.-).
 GN ANAS platyrhynchos (Domestic duck).
 OS Anas platyrhynchos (Domestic duck).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 CC Neognathae; Anseriformes; Anatidae; Anas.
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-FOREBRAIN;
 RA KIMURA N.; KUROSAWA N.; KONO K.; TSUKADA Y.;
 RA Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 RL -1 TISSUE SPECIFICITY: BRAIN.
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 -1 FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute has no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB009304; BAA23765.1;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 KM NP_BIND 48 NADP (BY SIMILARITY).
 FT ACT_SITE 201 201 BY SIMILARITY.
 FT SEQUENCE 312 AA: 33896 MW: 33896 B: 3DB3A6D CRC32:
 DH3_MOUSE Length: 312 February 14, 2000 08:02 Type: P Check: 9157

1 MIPAGLLMW VGALGALYAA VRGALGLIGA LRVWGIGAGR AALGPGLGN
 51 AVYATATGCI GKAYAKELAK RGMVVALISR SKKIDQVAG EITBQYVER
 101 KYIVADFGER EDIYDRIRAG LEGLEIGVL NNVGISYSYP EYFIDVDDL
 151 KTDKMININ IMVCKMTRL VLPGLMERSK GYVLINISSA GMYPTPLLT
 201 YSASKAFVDY FSRDLHNEY SKGIIVQSM PYVATKMSK ISKPSDKPT
 251 PETYVRAIG TVGLQSTING CIPHAFMGV FSILPTIYV NLMTKTKOI
 301 RARFLKRRK EK

11AA_SEQUENCE 1.0 STANDARD; PRT; 276 AA.
 ID DH3_HUMAN
 AC 075828;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE CARBONYL REDUCTASE [NADPH] 3 (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL REDUCTASE 3).
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98414514.
 RA WATANABE K., SUGAMARA C., ONO A., FUKUZUMI Y., ITAKURA S.,
 RA YAMAZAKI M., TASHIRO H., OSEOGAWA K., SOEDA E., NOMURA T.;
 RT "Mapping of a novel human carbonyl reductase, CBR3, and ribosomal
 RT pseudogenes to human chromosome 21q22.2.";
 RL Genomics 52:95-100(1998).
 CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB004854; BAA33500.1; -;
 DR EMBL: AB003151; BAA34207.1; -;
 DR MIM: 603608; -;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFM: PFO0106; adh_short; 1.
 KW Oxidoreductase; NADP.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT NP_BIND 9 33 NADP (BY SIMILARITY).
 FT ACT_SITE 193 193 BY SIMILARITY.
 SQ SEQUENCE 276 AA; 30719 MW; AETD716B CRC32;
 DHCA_HUMAN Length: 276 February 14, 2000 08:02 Type: P Check: 391 ..
 1 SSGSRVALVT GANRGIGLAI ARELCROFSG DVLTLARDAV RGAAYVOQLQ
 51 AEGISPRFHQ LDIDDLQSLR ALRDLRKRY GGLNLYVNA AVAKSDPM
 101 PFDIKAMTL KTNFATRNK CNELPIPMK HGRVNISSL QCLAFENCS
 151 EDQERFHS TLTEGLDVL MKRVEDTKN EVHERGWN SPYSGSLGV
 201 TVLSRIAR LDKRRADRI LVNCCGPGV KIDMGKXSI RTVBEGMTP
 251 VYALALPPDA TEPQGLVHD KYVQNM
 1AA_SEQUENCE 1.0 STANDARD; PRT; 276 AA.
 ID DHCA_HUMAN
 AC P16152;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CARBONYL REDUCTASE [NADPH] 1 (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
 DE REDUCTASE 1) (PROSTAGLANDIN-E2 9-REDUCTASE) (EC 1.1.1.189)
 DE (PROSTAGLANDIN 9-KETOREDUCTASE) (15-HYDROXYPROSTAGLANDIN DEHYDROGENASE
 DE (NADP+)) (EC 1.1.1.197).
 GN CBR1 OR CBR OR CRN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 89034082.
 RA WERNITZ B., BOHREN K.M., HEINEMANN G., VON WARTBURG J.-P.,
 RA GABBAY K.H.;
 RT "Human carbonyl reductase. Nucleotide sequence analysis of a cDNA and
 RT amino acid sequence of the encoded protein.";
 RL J. Biol. Chem. 263:16185-16188(1988).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-BREAST;
 RX MEDLINE: 90212644.

RA FORREST G.L., AKMAN S., KRUTZIK S., PAXTON R.J., SPARKES R.S.,
 RA DOROSHOW J., FELSTED R.L., MOHANDAS T., BACHUR N.R.;
 RT "Induction of a human carbonyl reductase gene located on chromosome
 RT 21.";
 RL Biochim. Biophys. Acta 1048:149-155(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92017676.
 RA FORREST G.L., AKMAN S., DOROSHOW J., RIVERA H., KAPLAN W.D.;
 RT "Genomic sequence and expression of a cloned human carbonyl reductase
 RT gene with daunorubicin reductase activity.";
 RL Mol. Pharmacol. 40:502-507(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98414514.
 RA WATANABE K., SUGAMARA C., ONO A., FUKUZUMI Y., ITAKURA S.,
 RA YAMAZAKI M., TASHIRO H., OSEOGAWA K., SOEDA E., NOMURA T.;
 RT "Mapping of a novel human carbonyl reductase, CBR3, and ribosomal
 RT pseudogenes to human chromosome 21q22.2.";
 RL Genomics 52:95-100(1998).
 RN [5]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE: 9313816.
 RA KROOK M., GHOSH D., STROEMBERG R., CARLQUIST M., JOERNVALL H.;
 RT "Carboxylethyllysine in a protein: native carbonyl reductase/NADP(+)-
 RT dependent prostaglandin dehydrogenase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:502-506(1993).
 CC -1- FUNCTION: CATALYZE THE REDUCTION OF A WIDE VARIETY OF CARBONYL
 CC COMPOUNDS INCLUDING THE ANTITUMOR ANTHRACYCLINE ANTIBIOTICS.
 CC CAN CONVERT PROSTAGLANDIN E2 TO PROSTAGLANDIN F2-ALPHA.
 CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.
 CC -1- CATALYTIC ACTIVITY: (5Z,13E)-(15S)-9,11,15-TRIHYDROXYPROSTA-5,13-
 CC DIENATE + NADP(+) = (5Z,13E)-(15S)-11,15-DIHYDROXY-9-OXOPROSTA-
 CC 5,13-DIENATE + NADPH.
 CC -1- CATALYTIC ACTIVITY: (5Z,13E)-(15S)-11-ALPHA,15-DIHYDROXY-9-
 CC OXOPROST-13-ENATE + NADP(+) = (5Z,13E)-11-ALPHA-HYDROXY-9,15-
 CC DIOXOPROST-13-ENATE + NADPH.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04056; AAA52070.1; -;
 DR EMBL: M62420; AAA17881.1; -;
 DR EMBL: AB003151; BAA33498.1; -;
 DR PIR: A31912; RDHUCB.
 DR PIR: S09013; S09013.
 DR HSP: P14061; 1PDM.
 DR MIM: 114830; -;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFM: PFO0106; adh_short; 1.
 KW Oxidoreductase; NADP; Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 1 1 ACETYLATION.
 FT NP_BIND 238 238 N6-(1-CARBOXYETHYL).
 FT ACT_SITE 9 33 NADP (BY SIMILARITY).
 FT ACT_SITE 193 193 BY SIMILARITY.
 SQ SEQUENCE 276 AA; 30244 MW; 470A0F07 CRC32;
 DHCA_HUMAN Length: 276 February 14, 2000 08:02 Type: P Check: 3202 ..
 1 SSGIHALVLT GGNKGIGLAI VRDLCLFSG DVLTLARDAVT RGAAYVOQLQ
 51 AEGISPRFHQ LDIDDLQSLR ALRDLRKRY GGLNLYVNA GIAKRVADPT

```

101 PPHIOAEVTM KTFNGTRDV CTELLPLIKR OGRVNVSSM MSVALKSCS
151 PELQCKFRSE TITEELVGL MKNFVEDTK GVHOGKQPS SAICVTIKIGV
201 TVLSRIHARK LSEORRGDKI LNVACCPGV RTDMAGPRAT KSPBEGAETP
251 VYIALPPDA EGPHQGFVQD KRYEOM

!!AA_SEQUENCE 1.0
ID DECA_MOUSE STANDARD: PRT: 276 AA.
AC P47727.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CARBONYL REDUCTASE [NADPH] (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
REDUCTASE).
GN CBR OR CBR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN:CBR/CJ; TISSUE:CBREBELLUM;
RA MEDLINE: 96299674.
RT Wei J., DLOUTH S.R., HARA A., GHETTI B., HODES M.E.;
RT Cloning a cDNA for carbonyl reductase (CBR) from mouse cerebellum:
RT murine genes that express cbr map to chromosomes 16 and 11."
RT Genomics 34:147-148(1996).
CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
FAMILY (SDR).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U31666; ARI9006.1; -
DR GDB: MGI:88284; CBR.
DR PROSITE: P500061; ADH_SHORT; 1.
DR PFM: PFM0106; adh_short; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT MOD_RES 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 9 NADP (BY SIMILARITY).
FT ACT_SITE 193 BY SIMILARITY.
FT SEQUENCE 276 AA; 30597 MW; 09438E74 CRC32;

DECA_MOUSE Length: 276 February 14, 2000 08:02 Type: P Check: 204

```

```

DE CARBONYL REDUCTASE [NADPH] (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
REDUCTASE).
GN CBR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-LIVER;
RA MEDLINE: 95197021.
RT GONZALES B., SARRA A., RIVERA H., KAPLAN W.D., YAM B., FORREST G.L.;
RT Cloning and expression of the cDNA encoding rabbit liver carbonyl
RT reductase."
RT Gene 154:297-298(1995).
CC -1- FUNCTION: CATALYZES THE REDUCTION OF A WIDE VARIETY OF CARBONYL
CC COMPOUNDS INCLUDING THE ANTITUMOR ANTHRACYCLINE ANTIBIOTICS.
CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: PRESENT IN LIVER AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
FAMILY (SDR).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U07051; AAA7670.1; -
DR EMBL: U09244; AAA82159.1; -
DR PROSITE: P500061; ADH_SHORT; 1.
DR PFM: PFM0106; adh_short; 1.
KW Oxidoreductase; NADP.
FT INIT_MET 0 BY SIMILARITY.
FT NP_BIND 9 NADP (BY SIMILARITY).
FT ACT_SITE 193 BY SIMILARITY.
FT VARIANT 23 A -> N (IN CLONE RCBR6).
FT VARIANT 79 V -> E (IN CLONE RCBR6).
FT VARIANT 91 A -> A (IN CLONE RCBR6).
FT VARIANT 214 H -> Q (IN CLONE RCBR6).
FT SEQUENCE 276 AA; 30521 MW; EAF92E81 CRC32;

DECA_MOUSE Length: 276 February 14, 2000 08:02 Type: P Check: 1426

```

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;
RX MEDLINE; 95220378.
RA WEIDMUTH B., MAEDER-HEINEMANN G., ERNST E.;
RT "Cloning and expression of cardonyl reductase from rat testis."
RL Eur. J. Biochem 228:473-478(1995).
CC -1- CATALYTIC ACTIVITY: R-COH-R' + NADP(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOL.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84349; CAA59088.1; -;
DR EMBL; X85886; CAA65230.1; -;
DR PROSITE; P50061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INITIATOR 1
FT MOTIFS 0
FT MOTIFS 1
FT ACT_SITE 193 193
FT ACT_SITE 30447 30447 MW; D9BC158B CRC32;
SQ SEQUENCE 276 AA; 2000 08:02 Type: P Check: 3202 ..
DHCA_RAT Length: 276 February 14, 2000 08:02
1 SSDREVALVT GANKGIFAI VDLCKRFLG DYVLTADES RGEANXKQD
51 TEGLSPRHQ LDIDNPOSIR ALRDLQLEY GGLVAVNNA GIAFKYVDT
101 PFHIAEYTM KTNFGTQDY CKELPIIKP QGRVAVNVS VSRALKSS
151 PELQKRFSE TITEELVGL MKFIEDAKK GVHAKGMPN SANGYKIKV
201 TVLSHIVARK LNERREDKI LNAACPGWV RTDAGKRAK KSEGEAEIP
251 VYLLALPQA BGRHGFQVD KVEPR
11AA SEQUENCE 1.0
ID P15127 STANDARD; PRT; 272 AA.
AC P15127
DT 01-FEB-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1996 (Rel. 34, Last annotation update)
DE GRANAICIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE 1
DE (EC 1.3.1.-) (ORF5)
OS Streptomyces violaceoruber.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J11511; CAA77599.1; -;
DR PIR; S25079; S25079.
DR HSSP; P18992; 2HSD.
DR PROSITE; P50061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 157 157
FT ACT_SITE 27080 27080 MW; 12DF5E53 CRC32;
SQ SEQUENCE 261 AA; 2000 08:02 Type: P Check: 3104 ..
DHKL_STRVN Length: 272 February 14, 2000 08:02
1 MTATATATA TGTAKAPVA LVGTATGIG LAIRRLAL GARTFLCARD
51 EERLAQTVKE LRGEGFDVG TVCDVADPAQ IAYVAAYQ RYGVVDILVN
101 NAGRSSGGAT AEIADLMMD VITINLTSF LMTKEVLNG GMLAKKRGRI
151 INIASIGGKQ GVHVAIPYA SKHGVGLK AGLELARTG ITVAACPGF
201 VEPMAERVR EHYAGIMQVS EEFTRDITN RYPLGRYVET REVANWEYL
251 VADDAAVTA QALVCGGIG NY
11AA SEQUENCE 1.0
ID DHKL_STRVN STANDARD; PRT; 261 AA.
AC P1177;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MONENSIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE (EC 1.3.1.-)
DE (ORF5)
OS Streptomyces cinamonensis.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11511; CAA77599.1; -;
DR PIR; S25079; S25079.
DR HSSP; P18992; 2HSD.
DR PROSITE; P50061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 157 157
FT ACT_SITE 27080 27080 MW; 12DF5E53 CRC32;
SQ SEQUENCE 261 AA; 2000 08:02 Type: P Check: 3104 ..

modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16300; CAA34368.1; -;
DR EMBL; X16144; CAA34263.1; -;
DR PIR; S05397; S05397.
DR HSSP; P03617; 11PK.
DR PROSITE; P50061; ADH_SHORT; 1.
DR PFAM; PF00678; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 168 168
FT ACT_SITE 28393 28393 MW; 1C3C79A3 CRC32;
SQ SEQUENCE 272 AA; 2000 08:02 Type: P Check: 3104 ..
DHKL_STRVN Length: 272 February 14, 2000 08:02
1 MTATATATA TGTAKAPVA LVGTATGIG LAIRRLAL GARTFLCARD
51 EERLAQTVKE LRGEGFDVG TVCDVADPAQ IAYVAAYQ RYGVVDILVN
101 NAGRSSGGAT AEIADLMMD VITINLTSF LMTKEVLNG GMLAKKRGRI
151 INIASIGGKQ GVHVAIPYA SKHGVGLK AGLELARTG ITVAACPGF
201 VEPMAERVR EHYAGIMQVS EEFTRDITN RYPLGRYVET REVANWEYL
251 VADDAAVTA QALVCGGIG NY
11AA SEQUENCE 1.0
ID DHKL_STRVN STANDARD; PRT; 261 AA.
AC P1177;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MONENSIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE (EC 1.3.1.-)
DE (ORF5)
OS Streptomyces cinamonensis.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11511; CAA77599.1; -;
DR PIR; S25079; S25079.
DR HSSP; P18992; 2HSD.
DR PROSITE; P50061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 157 157
FT ACT_SITE 27080 27080 MW; 12DF5E53 CRC32;
SQ SEQUENCE 261 AA; 2000 08:02 Type: P Check: 3104 ..

DHRR_STRCM Length: 261 February 14, 2000 08:02 Type: P Check: 4473

```

1 MROSTSRVAL VTGATSGIGL ATARLLAOG HLYELGARTE SVIATVVAL
51 RNDGLEAEOQ VLDVBDGASV TAFVQANADR YGRIDVLYNN AGRSGGVTVA
101 DLIDELMDV IDTNLSVFR MTRAVITTCG MTRERGRIT NASTAGKOG
151 VVLGAPYSAS KHVGVEFKA LGNELAPTCI TVNACGCVY ETPMAQRYAO
201 GYAAADTE EAILTKFOAK IPLGRSTPE EVAGLIGTVA SDTASITSO
251 ALWVCGGLGN F

11AA_SEQUENCE 1.0 STANDARD: PRT: 252 AA.
ID DTE_BACSU
AC P39577:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DTE PROTEIN.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
CC (1)
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RX MEDLINE; 95020537.
RA GLASER P., KONST F., ARNOLD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCUC M., LUBROCHINSKY B., MARCELINO L., MOSZER I.,
RA PRASECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DAVCHIN A.,
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 393 degrees."
RT Mol. Microbiol. 10:371-384(1993).
RN (2)
RN CHARACTERIZATION.
RP MEDLINE; 95318144.
RX PERGO M., GLASER P., MINUTELLO A., STRAUCH M.A., LEOPOLD K.,
RA FISCHER W.;
RT "Incorporation of D-alanine into lipoteichoic acid and wall teichoic
RT acid in Bacillus subtilis. Identification of genes and regulation.";
RT J. Biol. Chem. 270:15598-15606(1995).
CC -1- PATHWAY: BIOSYNTHESIS OF D-ALANYL-LIPOTEICHOIC ACID.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: TO 3-OXOACYL-LACYL-CARRIER PROTEIN) REDUCTASES.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73124; CA51557.1;
CC EMBL: Z59123; CAB18880.1;
CC PIR: S39656; S39656.
CC SUBTILIST: BG10547; DTE.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC PFAM: PF00106; adh_short; 1.
CC Oxidoreductase.
CC NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 152 152 BY SIMILARITY.
CC SEQUENCE 252 AA; 28266 MW; CC2ALAD CRC32;

DTE_BACSU Length: 252 February 14, 2000 08:02 Type: P Check: 2631
1 MKMTNVTYLI TGGSGAGIGLE LAKRLLEIGN EVIIGSRSEA KLAERKQOLP

```

```

51 NHTKOCVYA DRSOREALYE WALKREYNTLN VLVNNGIAOK EIDFKGTTEE
101 LEVDGDEIEL NFOAPYTLASA LETPHLMKOP EAAIVQVTSQ IAFNPLAVYP
151 VYCATKALAH FSFILTLLHHV ROTSVEVTEM APPAVDVLGN QKSRDKGLIT
201 YRGISSEERY QYFLDGLKEG KOETINERVE GLMDATRADY DLFQOMNTO
251 EN

11AA_SEQUENCE 1.0 STANDARD: PRT: 165 AA.
ID DSBF_VIBAL
AC Q56578;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
GN DSBF.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
CC (1)
RN SEQUENCE FROM N.A.
RP STRAIN-B138-2;
RX NAKAMURA T., ENOMOTO H., ENOMOTO T.;
RT Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS SUCH AS PHO A OR OMPA. ACTS BY OXIDIZING
CC THE DSBF PROTEIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D83728; BAA12087.1;
CC KMW Oxidoreductase;
CC FT DOMAIN 1 16
CC FT TRANSMEM 3 33 POTENTIAL.
CC FT TRANSMEM 17 51 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 34 51 POTENTIAL.
CC FT TRANSMEM 52 67 POTENTIAL.
CC FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 75 92 POTENTIAL.
CC FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
CC FT TRANSMEM 148 165 POTENTIAL.
CC FT NON_TER 165
CC FT SEQUENCE 165 AA; 18745 MW; FA018063 CRC32;

DSBF_VIBAL Length: 165 February 14, 2000 08:02 Type: P Check: 2736
1 MTNLNSLNF SKGRSLWLL LLEVFEFC ALYRQHWML APCVMCIYER
51 VYAMGCVAA IVGLMAPNRP IFRWGLLGN GLSSKGLLL AOHVHYOFPN
101 PSPATCDLF VTFPSMRPN QNAPWFEAY GDCKRIWQF IDLSMPQWLV
151 VIRAGNLIAL ALIVI

11AA_SEQUENCE 1.0 STANDARD: PRT: 367 AA.
ID DSI1_HUMAN
AC P28562;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
DE (MAP KINASE PHOSPHATASE-1) (MKP-1) (PROTEIN-TYROSINE PHOSPHATASE
DE CL100) (DUAL SPECIFICITY PROTEIN PHOSPHATASE HVH1).
GN DUSP1 OR PTPN10 OR MKP1 OR CL100 OR HVH1.

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX TISSUE-FORESKIN;
 RA MEDLINE: 93024952.
 RT KEYES S.M., EMSLIE E.A.;
 RL "Oxidative stress and heat shock induce a human gene encoding a
 protein-tyrosine phosphatase";
 Nature 359:644-647(1992).
 CC -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
 KINASE ERK2 ON BOTH THR-183 AND TYR-185.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- INDUCTION: BY OXIDATIVE STRESS AND HEAT SHOCK
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68277; CAA48338.1; -;
 DR PIR: S29090; S29090.
 DR HSSP: P51452; 1YHR.
 DR MW: 600714; -;
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PRAM: PF00581; Rhodanese; 1.
 DR PRAM: PF00782; DSPc; 1.
 KW Hydroxylase; Cell cycle;
 FT DOMAIN 175 367 CATALYTIC.
 FT ACT_SITE 258 258 BY SIMILARITY.
 FT SEQUENCE 367 AA; 39297 MW; DA333420 CRC32;
 DUSL_HUMAN Length: 367 February 14, 2000 08:02 Type: P Check: 5401 ..
 1 MYMEVGLTDA GGLRLALGER AAQCLLDRC SFPAFNAGHI AGSVNVRST
 51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVYLDERS AALGAKRDG
 101 TLALAAGALC REARAQVEF LKGGEAFSA SCEPLCSKOS TPKGLSLPLS
 151 TSVPSDAESG CSSCSTPLYD QGGVEILPF LYLSAYHAS RKMIDLALGI
 201 TALINVSANC PNHEGHYQ KSIPIVEDNHK ADISSWENEA IDFLDSIKNA
 251 GGRVFNCOA GISRSATICL AYLMRTNRYK LDEAFEVNQ RRSITSPNFS
 301 FMGQLLOFES QVLAHCSAE AGSPAMAVLD RGTSTTVEN FVSLPVPHT
 351 NSALNLYKSP ITTSPSC
 //AA_SEQUENCE 1.0
 ID DUSL_MOUSE STANDARD: PRT: 367 AA.
 AC P28563;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (MAP KINASE PHOSPHATASE-1) (MPK-1) (PROTEIN-TYROSINE PHOSPHATASE
 DE 3CH134) (PROTEIN-TYROSINE PHOSPHATASE ERP).
 GN DUSP1 OR PTPN10 OR MKP1 OR 3CH134 OR PTPN16.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RX MEDLINE: 92158357.
 RA CHARLES C.H., ABLER A.S., LAU L.F.;
 RT "CDNA sequence of a growth factor-inducible immediate early gene and
 RT characterization of its encoded protein";
 RL Oncogene 7:187-190(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX TISSUE-LIVER;
 RA MEDLINE: 93360956.
 RA NOGUCHI T., KETZ R., CHEN L., MATTER M.-G., CARRASCO D., BRAVO R.;
 RT "Structure, mapping, and expression of *erp*, a growth factor-inducible
 RT gene encoding a nontransmembrane protein tyrosine phosphatase, and
 RT effect of *erp* on cell growth";
 RL Mol. Cell. Biol. 13:5195-5205(1993).
 RN (3)
 RP CHARACTERIZATION
 RX MEDLINE: 94037096.
 RA SUN H., CHARLES C.H., LAU L.F., TONKS N.K.;
 RT "MRP-1 (3CH134), an immediate early gene product, is a dual
 RT specificity phosphatase that dephosphorylates MAP kinase in vivo";
 RL Cell 75:487-493(1993).
 CC -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
 KINASE ERK2 ON BOTH THR-183 AND TYR-185.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- INDUCTION: BY GROWTH FACTORS.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X61940; CAA43944.1; -;
 DR EMBL: S64851; AAB27882.1; -;
 DR PIR: S24411; S24411.
 DR HSSP: P51452; 1YHR.
 DR MGD: MG1:105120; PTPN16.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PRAM: PF00581; Rhodanese; 1.
 DR PRAM: PF00782; DSPc; 1.
 KW Hydroxylase; Cell cycle;
 FT DOMAIN 175 367 CATALYTIC.
 FT ACT_SITE 258 258
 FT MUTAGEN 258 258 C->S: LOSS OF ACTIVITY.
 FT SEQUENCE 367 AA; 39369 MW; 17B905B5 CRC32;
 DUSL_MOUSE Length: 367 February 14, 2000 08:02 Type: P Check: 5871 ..
 1 MYMEVGLTDA GGLRLALGER AAQCLLDRC SFPAFNAGHI AGSVNVRST
 51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVYLDERS ASDGAKRDG
 101 TLALAAGALC REARSTQVEF LKGGEAFSA SCEPLCSKOS TPKGLSLPLS
 151 TSVPSDAESG CSSCSTPLYD QGGVEILPF LYLSAYHAS RKMIDLALGI
 201 TALINVSANC PNHEGHYQ KSIPIVEDNHK ADISSWENEA IDFLDSIKNA
 251 GGRVFNCOA GISRSATICL AYLMRTNRYK LDEAFEVNQ RRSITSPNFS
 301 FMGQLLOFES QVLAHCSAE AGSPAMAVLD RGTSTTVEN FVSLPVPHT
 351 NSALNLYKSP ITTSPSC

```

11AA_SEQUENCE 1.0 STANDARD; PRT; 367 AA.
ID DUSL_RAT
AC 064623;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
DE (MAP KINASE PHOSPHATASE-1) (MFK-1) (PROTEIN-TYROSINE PHOSPHATASE
DE CL100).
GN DUSP1 OR CL100.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP STRAIN-SPRAGUE-DAWLEY; TISSUE-LUNG;
RC MUDA M., SCHLEGEL W., ARKINSTAL S.;
RA Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
CC -1- KINASE ERK2 ON BOTH THR-183 AND TYR-185.
CC -1- CATALYTIC ACTIVITY: PROTEIN-TYROSINE PHOSPHATASE + H(2)O =
CC PROTEIN-TYROSINE + ORTHOPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X84004; CA55828.1;
CC DR HSSP; P51452; TYR-PROTEIN PHOSPHATASE 1; 1.
CC DR PROSITE; PS00383; TYR-PROTEIN PHOSPHATASE 2; 1.
CC DR PROSITE; PS00383; TYR-PROTEIN PHOSPHATASE 2; 1.
CC DR PROSITE; PS00383; TYR-PROTEIN PHOSPHATASE 2; 1.
CC DR PFAM; PF00581; Rhodanese; 1.
CC DR PFAM; PF00782; DsPc; 1.
CC DR Hydrolase; Cell cycle.
CC FT DOMAIN 175 367 CATALYTIC.
CC FT ACT_SITE 258 367 BY SIMILARITY.
CC FT SEQUENCE 367 AA; 39541 MW; 9DE26E3 CRC32;
SQ
DUSL_RAT Length: 367 February 14, 2000 08:02 Type: P Check: 6443
1 MMEVEGIIDA GGLRALRER AAQCILLDCR SFAFNAGHI VGSVNVREIT
51 IYVRRAKGM GLEHIVPTE LRGRLAGAY HAVYLLDERS AALDGAKRDC
101 TLALAGALC REARSTQVF LOGGYEAFSA SCPELCSKOS TPGSLSLPIS
151 TSPDDEASEG CSSCSTPLYD OGGPVEILSF LYLGSAYHAS RKMDDALCI
201 TALINVSANC PNHEGHYQY KSIPEVDNHK ADISSWENEA IDEIDSICKA
251 GGRVPHQCA GISRSATICL AYLMRTNRYK IDEAFEFYKQ RRSIIIPNES
301 FMGOLLOFES QVLAPHCSAE AGSPAMAVLD RGTSTTVFN FVSPIVHPPT
351 NSALNTYOSP ITTSPSC
11AA_SEQUENCE 1.0 STANDARD; PRT; 355 AA.
ID DVRL_BRARE
AC P35621;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DVR-1 PROTEIN PRECURSOR.
DE VGI OR DVR-1.
GN Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

```

```

OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 94009920.
RA HELDE K.A., GRUNWALD D.J.;
RT The DVR-1 (vgr1) transcript of zebrafish is maternally supplied and
RT distributed throughout the embryo.
RL Dev. Biol. 135:418-426(1993)
CC -1- FUNCTION: SERVES TO FACILITATE THE DIFFERENTIATION OF EITHER
CC MESODERM OR ENDODERM EITHER AS A COFACTOR IN AN INSTRUCTIVE
CC SIGNAL OR BY PROVIDING PERMISSIVE ENVIRONMENT.
CC -1- SUBUNIT: HOMODIMER (PROBABLY).
CC -1- TISSUE SPECIFICITY: ABUNDANT IN OVARIES AND EGGS, AND EQUALLY
CC DISTRIBUTED AMONG ALL BLASTOMERES.
CC -1- DEVELOPMENTAL STAGE: CONCENTRATED IN THE STREAMING CYTOPLASM IN
CC JUST-FERTILIZED EGGS. EVENLY PARTITIONED DURING CLEAVAGE AMONG ALL
CC BLASTOMERES. ABSENT IN THE YOLK CELL DURING CLEAVAGE, BLASTULA AND
CC GASTRULA STAGES. DISTRIBUTED HOMOGENEUSLY AMONG ALL CELLS OF THE
CC GASTRULA.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00931; AAC27347.1;
CC DR HSSP; P18075; BMP.
CC DR ZFIN; ZDB-GENE-980526-389; VGI.
CC DR PROSITE; PS00250; TGF-BETA; 1.
CC DR PFAM; PF00019; TGF-beta; 1.
CC DR Growth factor; Mitogen; Glycoprotein; Signal.
CC KW SIGNAL 15 240
CC FT PROPEP 1 15 POTENTIAL.
CC FT CHAIN 241 355 DVR-1 PROTEIN.
CC FT DISULFID 254 320 BY SIMILARITY.
CC FT DISULFID 283 352 BY SIMILARITY.
CC FT DISULFID 287 319 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 319 319 POTENTIAL.
CC FT CARBOHYD 108 179 POTENTIAL.
CC FT CARBOHYD 179 296 POTENTIAL.
CC FT CARBOHYD 296 355 POTENTIAL.
CC FT SEQUENCE 355 AA; 40201 MW; B6B2B96C CRC32;
SQ
DVRL_BRARE Length: 355 February 14, 2000 08:02 Type: P Check: 2951
1 MFLVLRACL LTLISCSPAE DGLVOEKL FLSMGLMSRP KSHHAAYPS
51 OMKRIKQAS KQVNDPCV SEYVGNITV RFWDQGSLSI SAPAVHSFNC
101 VRKHLFFNNS VLEVEQLSL AOLMKFKQD LLLGLPHVFS VLIYRLKLT
151 LKGVTHSSR KLLQSOTLSP GAHASVLYNL TNLASWRRK EKNFGQLEL
201 QVHNLNMLH DHAYVOIPDI HATLVVVSIN PLQCHSRKR SASIYLPVP
251 SNVCKPRRLX IDFKDVGWD WIIAPOGYLA NYCHQCEPFP LSESLNGTNH
301 AILQTLVHSF DPKGTPOPC VPILKSPISW IYDNDNDV LRYEDNAVVD
351 ECGCR
11AA_SEQUENCE 1.0 STANDARD; PRT; 255 AA.
ID E182_DROME
AC P08761;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1988 (Rel. 09, Last annotation update)

```

DE ECDYSONE-INDUCED PROTEIN 28/29 KD.
GN EIP28/29.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
RN Ephydroidea; Drosophilidae; Drosophila.
1) SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE: 87060956.
RA CHERBAS L., SCHULTZ R.A., KOEHLER M.M.D., SAVAKIS C., CHERBAS P.;
RT "Structure of the EIP28/29 gene, an ecdysone-inducible gene from
RT Drosophila.";
RL J. Mol. Biol. 189:617-631(1986).
CC -1- INDUCTION: BY ECDYSONE.
CC -1- SIMILARITY: TO BACTERIAL AND EUKARYOTIC PEPTIDE METHIONINE
CC SULFOXIDE REDUCTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58286; CAA41223.1;
DR EMBL: X04024; CAA27657.1;
DR EMBL: X04024; CAA27658.1;
DR EMBL: X04521; CAA28205.1;
DR PIR: A24254; A24254.
DR FLYBASE: FBgn0000385; EIP71CD.
DR PFAM: PF01625; PMSR; 1.
KW Alternative splicing.
FT VASPLIC 79
FT CONFLICT 252 252 MISSING (IN ISOFORM EIP29).
FT SEQUENCE 255 AA; 28332 MW; 61E72253 CRC32;
EIP28_DROME Length: 255 February 14, 2000 08:02 Type: P Check: 68
1 MSLLTTSVT HPELKLSTV RNEQELNIS PVHDVATRA TATFNGCFW
51 GAESLYGATR GVLRITVGYA GGSDLPYTR KMGDHEVLE IDYDTVTSF
101 KELLDLFWNN HEYGLTPPK RQYASLLYH DEQKQVANA SKLEQERRA
151 PEITTEIAS KENFYPAEAY HOKYRLQGHK DLASSINLSP KLOTSTYAT
201 KLNGYLAGVG GIEQKAEAE TTGSDAHPPA VLLPRGAER GPGSLLTWP
251 NVHRR
1) AA-SEQUENCE 1.0
ID ELAC-TRIVU STANDARD; PRT: 102 AA.
AC Q29143;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 39, Last annotation update)
DE EARLY LACTATION PROTEIN PRECURSOR.
GN ELP
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Metatheria; Dipodomorpha; Phalangeridae; Trichosurus.
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-30.
RC TISSUE-MAMMARY GLAND;
RA FIORTE C.P., GRIGOR M.R.;
RT "A novel mammary protein expressed by the mammary gland only during
RT the early lactation and related to the Kunitz proteinase inhibitors.";
RL Arch. Biochem. Biophys. 330:59-64(1996).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY DURING EARLY LACTATION PHASE.
CC -1- TISSUE SPECIFICITY: EXPRESSED BY THE MAMMARY GLAND.

CC -1- PTM: N-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34208; AAB08977.1;
DR HSSP: P00974; IPT.
DR PROSITE: PS00280; BPTI_KUNITZ; 1.
DR PFAM: PF00014; Kunitz_BPTI; 1.
KW Serine protease inhibitor; Signal; Lactation.
FT SIGNAL 1 20
FT CHAIN 21 102
FT DISULFD 43 92
FT DISULFD 52 76
FT DISULFD 68 89
FT ACT SITE 53 34
FT CARBOHYD 34 34
FT CARBOHYD 62 62
FT SEQUENCE 102 AA; 11407 MW; D7983676 CRC32;
ELAC-TRIVU Length: 102 February 14, 2000 08:02 Type: P Check: 5524
1 MKFTIALCL ASLVGMTSS EKLDRIRAN SLENLSRVV SLCLPSGRG
51 NCDSQLIRYF YNATSHCEY FLYSGCNGNG NNEDSLCCCL KTCRLNKRYN
101 NN
1) AA-SEQUENCE 1.0
ID ERP2-YEAST STANDARD; PRT: 215 AA.
AC P39704;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ERP2 PROTEIN PRECURSOR.
GN ERP2 OR YAL007C OR PUN54.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE: 95026152;
RA CLARK M.W., KENG T., STORMS R.K., ZHONG W., FORTIN N., ZENG B.,
RA DELANEY S., CHELLETTE B.F.F., BARTON A.B., KABACK D.B., BUSSEY H.;
RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
RT the 42 kbp SPO7-CENT-CO115 region.";
RL Yeast 10:535-541(1994).
RN (2)
RP CHARACTERIZATION.
RA MARICHO M., HENTHORN D.C., WILSON R., SOLARI R., THOMAS D.,
RA BERGERON J., ROWLEY A.;
RL Unpublished observations (JUL-1998).
CC -1- FUNCTION: INVOLVED IN VESICULAR PROTEIN TRAFFICKING.
CC -1- SUBUNIT: ASSOCIATES WITH ERP24 AND ERP25.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENOPLASMIC
CC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ERP24/GRP25L FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: L22015; AAC04951.1; -
 DR PIR: S43448; S43448.
 DR PIR: S40896; S40896.
 DR SGD: L0004679; ERP2.
 DR PIR: P01105; EMP24_GP25L; 1.
 DR Transport; Protein transport; transmembrane; signal;
 KW Endoplasmic reticulum.
 FT SIGNAL 1 25
 FT CHAIN 26 215
 FT DOMAIN 26 182
 FT TRANSMEM 183 203
 FT DOMAIN 204 215
 FT SEQUENCE 215 AA; 24063 MW; 50CD1406 CRC32;
 SO
 EMP2_YEAST Length: 215 February 14, 2000 08:02 Type: P Check: 5386 ..

1 MIKRIALPS FIYILIALY NSVAASSYA PVAISAPFS KECLYDMVT
 51 EDDSLAVGYQ VLTGNFEID EDITAPDSV ITSEKOKYS DELKSEGVG
 101 KTFPCFSNNY GTALKVEIT LEKEKLIDE HEADVNNNDI IANNAVEED
 151 RNLNITL NTLAREWEN NSTVNSTESR LTMILILI IYVISAQV
 201 LLIQFLFTR GKNV
 11AA_SEQUENCE 1.0 STANDARD; PRT; 270 AA.
 ID ESI_2BRARE
 AC 090237;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ESI PROTEIN PRECURSOR.
 GN ESI.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygia; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprininae; Rasbora; Danio.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EYE;
 RX MEDLINE: 97410119.
 RA CHANG H.; GILBERT W.;
 RT "A novel zebrafish gene expressed specifically in the photoreceptor
 cells of the retina."
 RL Biochem. Biophys. Res. Commun. 237:84-89 (1997).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN PHOTORECEPTOR CELLS
 OF THE RETINA.
 CC -1- SIMILARITY: BELONGS TO THE ESI FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: U10403; AAC60261.1; -
 DR ZFIN: ZDB-GENE-980526-186; ESI.
 KW Mitochondrion; Transmembrane; Transmembrane protein;
 FT TRANSIT 1 270
 FT CHAIN 270
 FT SEQUENCE 270 AA; 30685 MW; 4B65565D CRC32;
 SO
 ESI_2BRARE Length: 270 February 14, 2000 08:02 Type: P Check: 1627 ..

1 MASRALIAK QAAALVROP ACLMHGGDW GNMGNINIAV VSGCGWMDG
 51 TDHEAAYTM YHLSRNGARF QIFAPNOOM HVDHMKOP SSSDNINIM
 101 ESARFSGOG MMQNDLSKL DANSDAVIF PGHGIYKMM STFSKDGKDC

151 KLNDDVERVL KDFHARKPI GLSSNAPILLA CRVLPSEVI MGVERDESSR
 201 WGRPENTINNY QAVKSMGARH NTRPEPEAVY DEKKVISTP TFWMDYHY
 251 HYIPDIGNM VAVHVMKATK
 11AA_SEQUENCE 1.0 STANDARD; PRT; 318 AA.
 ID ETPA_MYCLE
 AC 033096;
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETP) (ELECTRON
 DE TRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETPLS).
 GN ETPA OR FIXB OR MCB637.04.
 OS Mycobacterium leprae.
 GN Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA OLIVER K.; HARRIS D.; PARKHILL J.; BARRELL B.G.; RAJANDREAM M.A.;
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
 ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE
 ELECTRON TO THE MAIN RESPIRATORY CHAIN VIA ETP-UBIQUINONE
 OXYDOREDUCTASE (ETP DEHYDROGENASE) (BY SIMILARITY).
 CC -1- COFACTOR: CONTAINS ONE MOLECULE OF FAD PER DIMER (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -1- SIMILARITY: BELONGS TO THE ETP ALPHA-SUBUNIT / FIXB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: Z99263; CAB16419.1;
 DR PROSITE: PS00696; ETP_ALPHA; 1.
 DR PIR: P00766; ETP_Alpha; 1.
 KW Electron transport; Flavoprotein; FAD (ADP PART) (POTENTIAL).
 FT NE_BIND 257 285
 FT SEQUENCE 318 AA; 31837 MW; 9CF35B35 CRC32;
 SO
 ETPA_MYCLE Length: 318 February 14, 2000 08:02 Type: P Check: 3764 ..

1 MAELVYEH TEGALKVSA ELITAAVYG EPAVVGTP GTSAPLYDGL
 51 KTAGAKIYV AEDADAKYL ITFVVDVLA LAESSAPAV LLAATDGKE
 101 IGGRLARIG SGLVDVVDV REGAVGVSHV EGVGEIVEAO ANGDTPIYIV
 151 RAGAVEAQA EGAGGVQSVV VPAPAEVATK ITRAPAVVD NRPDLTEAVY
 201 VVSGRGVGS ADNEVSEAL ADSIGAVGA SRAAVDSGY PGFOIISQTS
 251 KTVSPOLYIA LGISGALQHR AGMOTSKIY AVNKDEEADI FEIADFGVVG
 301 DLFKVAPOLT DGIKARKG
 11AA_SEQUENCE 1.0 STANDARD; PRT; 144 AA.
 ID EXBD_NEIGO
 AC 006434;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE BIOPOLYMER TRANSPORT EXBD PROTEIN.
 GN EXBD.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-FA19;
RX MEDLINE; 97285757.
RA BISMAS G.D., ANDERSON J.E., SPARKLING P.F.;
RT "Cloning and functional characterization of *Neisseria gonorrhoeae*
tomb, exb and exd genes."
RL MOJ. Microbiol. 24:169-179(1997).
CC -1- FUNCTION: INVOLVED IN THE TONB-DEPENDENT ENERGY-DEPENDENT
CC TRANSPORT OF VARIOUS RECEPTOR-BOUND SUBSTRATES (BY SIMILARITY).
CC -1- SUBUNIT: THE ACCESSORY PROTEINS EXB AND EXD SEEM TO FORM A
CC COMPLEX WITH TONB (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE EXBD / TOLR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U77738; AAC44836.1; -

EXBD_NEIMO length: 144 February 14, 2000 08:02 Type: P Check: 4894 ..

1 MAFGSMNSD DSPMSDINVT PLVDVVLVLL IVEITMPVL TSHIPLEPT
51 ASEQNKKDK QPKDPLRLTI DANGGYVYG DSASKVEIGE VESRLKAKE
101 QNENVIYVIA ADKAVEYDYV NNALEAARQA GITKIGVTE TKAO

11AA_SEQUENCE 1.0
ID EXBD_NEIME STANDARD; PRT; 144 AA.
AC P25716;
DT 13-DEC-1998 (Rel. 37, Created)
DT 13-DEC-1998 (Rel. 37, Last sequence update)
DT 13-DEC-1998 (Rel. 37, Last annotation update)
DE BIOPOLYMER TRANSPORT EXBD PROTEIN.
GN EXBD.
OS *Neisseria meningitidis*.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROGROUP C;
RX MEDLINE; 97158676.
RA STOJILJKOVIC I., SRINIVASAN N.;
RT "Neisseria meningitidis tomb, exb, and exd genes: Ton-dependent
utilization of protein-bound iron in *Neisseriae*."
RL J. Bacteriol. 179:805-812(1997).
CC -1- FUNCTION: INVOLVED IN THE TONB-DEPENDENT ENERGY-DEPENDENT
CC TRANSPORT OF VARIOUS RECEPTOR-BOUND SUBSTRATES (BY SIMILARITY).
CC -1- SUBUNIT: THE ACCESSORY PROTEINS EXB AND EXD SEEM TO FORM A
CC COMPLEX WITH TONB (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE EXBD / TOLR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U77738; AAC44836.1; -

KW Transport; Protein transport; Transmembrane; Inner membrane.
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 39 POTENTIAL.
FT DOMAIN 40 144 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 144 AA; 15514 MW; 6347P02B CRC32.

EXBD_NEIME length: 144 February 14, 2000 08:02 Type: P Check: 5831 ..

1 MAFGSMNSD DSPMSDINVT PLVDVVLVLL IVEITMPVL TSHIPLEPT
51 ASEQNKKDK QPKDPLRLTI DANGGYVYG DSASKVEIGE VESRLKAKE
101 QNENVIYVIA ADKAVEYDYV NNALEAARQA GITKIGVTE TKAO

11AA_SEQUENCE 1.0
ID FABG_ECOLI STANDARD; PRT; 244 AA.
AC P25716; P78221;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 3-OXOACYL-(ACYL-CARRIER PROTEIN) REDUCTASE (EC 1.1.1.100) (3-RETACYL-
DE ACYL CARRIER PROTEIN REDUCTASE).
GN FABG.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC *Escherichia*.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 92210530.
RA RAWLINGS M., CROMAN J.E. JR.;
RT "The gene encoding *Escherichia coli* acyl carrier protein lies within
a cluster of fatty acid biosynthetic genes."
RL J. Biol. Chem. 267:5751-5754(1992).
CC (2)
CC SEQUENCE FROM N.A.
CC STRAIN-K12; MG1655;
CC MEDLINE; 97426617.
CC BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
CC RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
CC GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
CC MAU B., SHAO Y.;
CC "The complete genome sequence of *Escherichia coli* K-12."
CC Science 277:1453-1474(1997).
CC [3]
CC SEQUENCE FROM N.A.
CC STRAIN-K12;
CC MEDLINE; 97061202.
CC OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
CC IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
CC KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
CC MORI H., MOTOKURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
CC SAMPEI G., SEKI Y., TAKAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
CC YANO M., HORICHI T.;
CC "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."
CC DNA Res. 3:117-151(1996).
CC -1- CATALYTIC ACTIVITY (3R)-3-HYDROXYACYL-(ACYL-CARRIER PROTEIN) +
CC NADP(4-) -> 3-OXOACYL-(ACYL-CARRIER PROTEIN) + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84991; AAA23739.1; -

DR EMBL: AE000210; AAC74177.1; --
 DR EMBL: D90745; BAA55901.1; --
 DR PIR: B42147; B42147.
 DR HSSP: P19992; 2HSD.
 DR ECOSGENE: EG11318; FABG.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF00678; adh_short_c2; 1.
 KM Fatty acid biosynthesis; Oxidoreductase; NADP.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 151 151 BY SIMILARITY.
 FT CONFLICT 30 30 A -> G (IN REF. 1).
 SQ SEQUENCE 244 AA; 25560 MW; 07319E62 CRC32;
 FABG_ECOLI Length: 244 February 14, 2000 08:02 Type: P Check: 3492 ..
 1 MNEGRALV TGASRGIGRA IAEITLARGA KYIGTAISEN GAQALSDYLG
 51 ANGKGLMNV TDRASIESVL EKIRAEQEV DLYVNNAGIT RDNLNMRKD
 101 EEMNDIIEIN LSVFRLSKA YMAAMKRRH GRIITGIV GTMNGGQAN
 151 YAAAKGLIG FSKSLAREVA SRGITVNVVA PGFIETDMTR ALSDDQAGI
 201 LAQVPAGRLG GAQETANAVA FLASDEAYVI TSETLVHNG MYMV
 11AA_SEQUENCE 1.0
 ID FABG_HAEIN STANDARD: PRT: 242 AA.
 AC P43713;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
 DE ACYL CARRIER PROTEIN REDUCTASE).
 GN FABG OR H10155.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE: 95350630.
 RA FREISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KRELVAGE A.R., BRIT C.J., TOMS J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA KRENNER K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOW E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SADER D.M., BRANDON R.C.,
 RA FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEORGEAN N.S.M.,
 RA GHEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U37301; AAC21824.1; --
 DR HSSP: P19992; 2HSD.
 DR TIGR: H10155;
 DR PROSITE: PS00061; ADH_SHORT; 1.

DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF00678; adh_short_c2; 1.
 KM Fatty acid biosynthesis; Oxidoreductase; NADP.
 FT NP_BIND 8 32 NADP (BY SIMILARITY).
 FT ACT_SITE 149 149 BY SIMILARITY.
 SQ SEQUENCE 242 AA; 25507 MW; CBFCADF9 CRC32;
 FABG_HAEIN Length: 242 February 14, 2000 08:02 Type: P Check: 4159 ..
 1 MGKIALVTG STRIGIGRAA ELSKGAIV IGTAISEKA EASAYLGDK
 51 GKGLIANTVD KESIEITLLEQ IKNDGDDI LNVNAGITRD NLNMRKDE
 101 WEDIWOTNLT SVYHLKRAMI RSMKRRFRG IINIGSVGS TGNPGQTVYC
 151 AAKAGVGFSS KSLAKEVAAR GITVNVVAPG FIATDMTEVL TDEKAGILS
 201 NVPAGRLGEA KDIAKAVAFI ASDDAGYITG TTIHVNGLY LS
 11AA_SEQUENCE 1.0
 ID FABG_PSEAE STANDARD: PRT: 247 AA.
 AC O54438;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
 DE ACYL CARRIER PROTEIN REDUCTASE).
 GN FABG.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 OC Pseudomonas.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAOI;
 RA KUTCHMA A.J., HOANG T.T., SCHWEIZER H.P.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U91631; AAB94395.1;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF00678; adh_short_c2; 1.
 KM Fatty acid biosynthesis; Oxidoreductase; NADP.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 154 154 BY SIMILARITY.
 SQ SEQUENCE 247 AA; 25585 MW; C5B95CE0 CRC32;
 FABG_PSEAE Length: 247 February 14, 2000 08:02 Type: P Check: 9922 ..
 1 MSIGKVAVLV TGASRGIGOA IALEIGRLGA VVIGTATSAS GAETIAETLK
 51 ANGVEGAGLV LDVSSDESVA ATLEHIOHL GQPLIVVNA GITRDLNLR
 101 MDDEMFQV NTNLSLYRL SKAVLGMTR ARMGRINIG SVYGANGNG
 151 QNVYAAKAG LEFTRLAR EVGSRATVAV AVAGGFIQTD MRELPLEAQR
 201 EALIGQIPLG RLGADEIAR VVGFLASDGA AVYGVATVPV NGGMVMS
 11AA_SEQUENCE 1.0

```

ID  FABG_VIBHA  STANDARD;  PRT;  244 AA.
AC  P55336;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
GN  ACYL CARRIER PROTEIN REDUCTASE).
OS  FABG.
OS  Vibrionaceae.
OS  Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrion.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN: 96134997.
RX  MEDLINE: 96134997.
RA  SHEN Z., BYERS D.M.;
RT  "Isolation of Vibrio harveyi acyl carrier protein and the fabg, acpp,
RL  and fad genes involved in fatty acid biosynthesis."
CC  J. Bacteriol. 178:571-573(1996).
CC  -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC  NADP(+) -> 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC  -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC  PATHWAY.
CC  -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC  FAMILY (SDR).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: U39441; AAC43589.1;
DR  HSSP: P19992; 2HSD.
DR  PROSITE: PS00061; ADH_SHORT: 1.
DR  PFAM: PF00106; adh_short: 1.
DR  PFAM: PF00678; adh_short_C2: 1.
KW  Fatty acid biosynthesis; Oxidoreductase; NADP.
FT  NP BIND 10 34 NADP (BY SIMILARITY).
FT  ACT SITE 151 151 BY SIMILARITY.
FT  SEQUENCE 244 AA; 25519 MW; AE81EDC CRC32;
FABG_VIBHA Length: 244 February 14, 2000 08:02 Type: P Check: 5984 ..

1  MNEGLIALV TGASRGIGRA IAEILLVERGA TVIGTATSEG GAAAISEYLG
51  ENGKGLALNV TDVESIENTL KTIIDECGAI DILVNNAGIT RDNLNMRMD
101  DEMNDIINTN LPIYRMSKA VLGNMKKRA GRINWGSVY GTGNAGQTN
151  YAAKAGVIG FTKSMAREVA SRGVYNTVA PGFIETDWTX ALMDDDORAT
201  LSNVPAGRLG DPREIASAVV FLASPEAVYI TGETLHVNG MYWV

!!AA_SEQUENCE 1.0
ID  FENR_ECOLI  STANDARD;  PRT;  247 AA.
AC  P28861; P11007;
DT  01-JUL-1989 (Rel. 11, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  FERREDOXIN-NADP REDUCTASE (EC 1.18.1.2) (FNR) (FLAVODOXIN REDUCTASE)
DE  (FNR) (FNR) (METHYL VIIOLOGEN RESISTANCE PROTEIN A) (DRI).
GN  FPR OR MYR.
OS  Escherichia coli.
OS  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RX  STRAIN-K12 / C600;
RX  MEDLINE: 93194782.
RA  BIANCHI V., REICHARD P., ELIASSON R., PONTIS E., KROOK M.,
RA  JOERNVALL H., HAGGARD-LJUNGQUIST E.;

```

"Escherichia coli ferredoxin NADP+ reductase: activation of E. coli anaerobic ribonucleotide reduction, cloning of the gene (fpr), and overexpression of the protein".

J. Bacteriol. 175:1590-1595(1993).

[2] SEQUENCE FROM N.A.

STRAIN-K12 / MG1655;

REMARK: 93347969

PLUNKETT G., III, BURLAND V.D., DANIELS D.L., BLATTNER F.R.;

"Analysis of the Escherichia coli genome. III. DNA sequence of the

region from 87.2 to 89.2 minutes."

Nucleic Acids Res. 21:3391-3398(1993).

[3] PRELIMINARY SEQUENCE OF 1-134 FROM N.A.

STRAIN-K12 / MC4100;

REMARK: 93015762.

TRONIGER V., BOOS W., SWEET G.;

"Molecular analysis of the glpFK regions of Escherichia coli and

Shigella flexneri."

J. Bacteriol. 174:6981-6991(1992).

[4] PRELIMINARY SEQUENCE OF 1-128 FROM N.A.

STRAIN-K12;

MEDLINE: 88198002.

MORIMOTO M.;

"Isolation and characterization of methyl viologen-sensitive mutants

of Escherichia coli K-12."

J. Bacteriol. 170:2136-2142(1988).

[5] SEQUENCE OF 1-13.

MEDLINE: 95050480.

JENKINS C.M., WATERMAN M.R.;

"Flavodoxin and NADPH-flavodoxin reductase from Escherichia coli

support bovine cytochrome P450C17 hydroxylase activities."

J. Biol. Chem. 269:27401-27408(1994).

[6] X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).

MEDLINE: 97293095.

INELMAN M., BIANCHI V., EKUND H.;

"The three-dimensional structure of flavodoxin reductase from

Escherichia coli at 1.7-A resolution."

J. Mol. Biol. 268:147-157(1997).

-1- FUNCTION: TOGETHER WITH FLAVODOXIN IS INVOLVED IN THE REDUCTIVE

ACTIVATION OF COBALAMIN-INDEPENDENT METHIONINE SYNTHASE, PYRUVATE

FORMATE LYASE AND ANAEROBIC RIBONUCLEOTIDE REDUCTASE. ALSO

PROTECTS AGAINST SUPEROXIDE RADICALS DUE TO METHYL VIIOLOGEN IN THE

PRESENCE OF OXYGEN.

-1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) -> OXIDIZED

FERREDOXIN + NADPH.

-1- COFACTOR: FAD.

-1- SUBUNIT: MONOMER.

-1- SIMILARITY: TO OTHER FERREDOXIN NADP REDUCTASES.

-1- CAUTION: REF.3 SEQUENCE WAS INCORRECT DUE TO A FRAMESHIFT.

REF.4 AUTHORS INCORRECTLY ASSIGNED TO BE PART OF FPR, THE C-

TERMINAL OF GLPX.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).


```

DR PDB: 1PDR: 17-SEP-97.
DR ECOGENE: EGI1518; FPR.
DR PFAM: PF00175; oxidored_fad: 1.
KM Oxidoreductase; Flavoprotein; NADP; FAD; 3D-structure.
FT INIT_MET 0 0
FT NP_BIND 49 75 FAD.
FT NP_BIND 109 125 NADP (RIBOSE PART) (BY SIMILARITY).
SQ SEQUENCE 247 AA; 27620 MW; 58D558A CRC32;

FENR_ECOLI Length: 247 February 14, 2000 08:02 Type: P Check: 3071 ..

1 ADVATGKVK VQNTDALS LTVHAPLPF TAGQETKGL EIDGERORA
51 YSVNSPDNP DLEFLYVVP DGKISPLLA LKPGDEVQV SEAGFVLD
101 EVHCEETLM LATGTATGPY LSLIQGLKD DFKNLVYVH AARADLST
151 LPMQELERK YEGKLRIOY VSRETAGSL TGRIPALIES GELESTIGLP
201 MKETSHVWL CGNPQVMDT QQLIKETROM TKHLRRPQH MTAEHYW

11AA_SEQUENCE 1.0 STANDARD: PRT: 135 AA.
ID FENR_SHIFL
AC P28901;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FERREDOXIN-NADP REDUCTASE (EC 1.18.1.2) (FNR) (FLAVODOXIN REDUCTASE)
DE (FLXR) (METHYL VIOLOGEN RESISTANCE PROTEIN A) (FRAGMENT).
GN FPR OR MYRA.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M4243;
RX MEDLINE: 93015762.
RA TUNINGER V., BOOS W., SWEET G.;
RT "Molecular analysis of the glpFKX regions of Escherichia coli and
RT Shigella flexneri."
RL J.Bacteriol. 174:6981-6991(1992).
CC -1- FUNCTION: TOGETHER WITH FLAVODOXIN IS INVOLVED IN THE REDUCTIVE
CC ACTIVATION OF COBALAMIN-INDEPENDENT METHIONINE SYNTHASE. PYRUVATE
CC FORMATE LYASE AND ANAEROBIC RIBONUCLEOTIDE REDUCTASE. ALSO
CC PROTECTS AGAINST SUPEROXIDE RADICALS DUE TO METHYL VIOLOGEN IN THE
CC PRESENCE OF OXYGEN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) -> OXIDIZED
CC FERREDOXIN + NADPH.
CC -1- COFACTOR: FAD.
CC -1- SIMILARITY: TO OTHER FERREDOXIN NADP REDUCTASES.
CC -1- CAUTION: THE SEQUENCE SHOWN BELOW WAS CORRECTED TO BE IN AGREEMENT
CC WITH THE CORRESPONDING E. COLI SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: Z11766; CAAT7813.1; ALT_SEQ.
DR PIR: S23906; S23906.
DR HSSP: P28861; 1FDR.
KW Oxidoreductase; Flavoprotein; NADP; FAD.
FT NP_BIND 50 76 FAD (BY SIMILARITY).
FT NP_BIND 110 125 NADP (RIBOSE PART) (BY SIMILARITY).
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 14878 MW; F6A7ADSD CRC32;

FENR_SHIFL Length: 135 February 14, 2000 08:02 Type: P Check: 1251 ..

1 MADWYTGAVT KYQNTWDALE SLTVHAPVLP FTAGQETKGL LEIDSERVQR

11AA_SEQUENCE 1.0 STANDARD: PRT: 135 AA.
ID FENR_SHIFL
AC P28901;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FIBRINOLASE (EC 3.4.24.72) (FIBRINOLYTIC PROTEINASE).
DE Agkistrodon contortrix conortrix (Southern copperhead).
DE Agkistrodon contortrix conortrix (Southern copperhead).
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosaurs;
DE Squamata; Scleroglossa; Serpentes; Colubroidae; Viperidae; Crotalinae;
OC Agkistrodon.
RN [1]
RP SEQUENCE.
RC TISSUE-VENOM.
RX MEDLINE: 93278288.
RA RANDOLPH A.S. JR., MASTAR F.R.;
RA MARKLAND F.S. JR., MASTAR F.R.;
RT "Amino acid sequence of fibrinolase, a direct-acting fibrinolytic
RT enzyme from Agkistrodon contortrix conortrix venom."
RT Protein Sci. 1:350-360(1992).
CC -1- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC ACTS IN HEMORRHAGE. IT CLEAVES FIBRINOGEN,
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF 14-ALA-1-LEU-15 IN INSULIN B CHAIN
CC AND CLEAVAGE OF 413-LYS-1-LEU-414 IN ALPHA CHAIN OF FIBRINOGEN.
CC -1- COFACTOR: BINDS ONE ZINC ION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPTOLYSIN SUBFAMILY.
CC PIR: A41827; HYSNFA.
DR HSSP: P34179; 2AIG.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PFAM: PF01421; Reptolysin; 1.
KW Hydroxylase; Metalloprotease; Zinc; Venom.
FT MOD_RES 1 1
FT VARIANT 1 1
FT METAL 143 143
FT ACT_SITE 144 144
FT METAL 147 147
FT METAL 153 153
FT METAL 158 158
FT DISULFID 160 165
FT DISULFID 160 182
FT VARIANT 189 189
FT VARIANT 192 192
SQ SEQUENCE 203 AA; 22908 MW; DC49D0E9 CRC32;

FIBR_AGCKO Length: 203 February 14, 2000 08:02 Type: P Check: 7566 ..

1 QORFPORYQV LYIYADHRAN TKYNGDSKI RQWHAIVNT INITYPLNI
51 QFTVLELEW SNQDLITVTS VSHDITLASEG NWRETDILRR QORDNQLLT
101 AIDFDGDTVG LAYVGMQQL KSTGYIQDH SAINLVALT MAHELCHNIG
151 MNHNGQCNC GANSCVNAAM LSDQPSKLP S DSKNDYQTF LTVNPDQIL
201 NKP

11AA_SEQUENCE 1.0 STANDARD: PRT: 335 AA.
ID FIBR_AGCKO
AC P37925;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FIBRIN PROTEIN PRECURSOR.
GN FIBR.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.

```

```

RN [1]
RP SEQUENCE FROM N.A.
RL SWENSON D.L., CLEGG S.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
CC ADHESION OF TYPE 1 FIBRINAE (BUT NOT NECESSARY FOR THE PRODUCTION
CC OF FIBRINAE). SEEMS TO BE MANNOSE BINDING ADHESIN.
CC -1- SIMILARITY: BELONGS TO THE FIMH / LPED FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L19338; AAA75420.1; -
DR STYGENE: SG10507; FIMH.
DR Fimbrin; Signal. 22
FT SIGNAL 1
FT CHAIN 23
SQ SEQUENCE 335 AA; 36056 MW; 2C61967B CRC32;
FIMH_SALTY Length: 335 February 14, 2000 08:02 Type: P Check: 6074

1 MKTISALLLA GTALFFTHPA LATVCRNSNG TADIDFYDLS DVFTSGNNOP
51 GOVVTLPERS AMGVNATCP AGTVNNTYR SYVELPYOS TEGNFKYKTL
101 NDYLLGAMST TDSVAGVSP PRNYILMGVD YNVSQKPFQ VODSKLVFKL
151 KVRPELNMV TIPQTFMTV YVTSTGDAL STPYTISYS GKVEVPQNC
201 VNAGVVEED FGDIGASLFS QAGAGNPPQ VTPQRTIAI KCTNVAQAY
251 LSNRLAEKRA SGOAMVSDNP DLGFVYVNSN GPTLPNNLS SKIPFHLDDN
301 AARVGRIRAW PISVTGIRKA EGPFTARGYL RVDYD
11AA_SEQUENCE 1.0
ID FLIH_BACSU STANDARD; PRT; 208 AA.
AC P23449;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE FLAGELLAR ASSEMBLY PROTEIN FLIH.
GN FLIH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN: 681258343.
RX ALBERTINI A.M., CRABB W.D., SCOFFONE F., GALIZZI A.;
RX The flia locus of Bacillus subtilis is part of a large operon coding
RX for flagellar structures, motility functions, and an AtPase-like
RX polypeptide.
RL J. Bacteriol. 173:3573-3579(1991).
CC -1- SIMILARITY: TO FLIH IN OTHER BACTERIA AND ALSO TO FLIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56049; CA39522.1; -
DR EMBL: 299112; CAB13496.1; -

```

```

DR PIR: S14496; S14496.
DR PIR: C42365; C42365.
DR SUBMITLIST: BG10242; FLIH.
DR Flagellin.
SQ SEQUENCE 208 AA; 23788 MW; 448BA6EA CRC32;
FLIH_BACSU Length: 208 February 14, 2000 08:02 Type: P Check: 1150

1 MARVEEEDR ISQANSHE NIRQIEQER NDMAERKRL IEKARQFE
51 OGVALGRKAE MKOYELLIQ ANTTMSRK AYEDKLEDAN BEIYELVAL
101 AKRWQOKSD DKEAFLLVQ QVINEVEYD DISIYDPY YETIFQORDE
151 IQQLLYECCR LGIYADEKAO KGTCTYETPF GRVDAVDTO LMQLKRLT
201 ALEAGAAE
11AA_SEQUENCE 1.0
ID FOS_AVINK STANDARD; PRT; 322 AA.
AC P23050;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE P55-V-FOS TRANSFORMING PROTEIN.
GN V-FOS.
OS Avian retrovirus NK24.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 88062820.
RX NISHIZAWA M., GOTO N., KAMAI S.;
RX "An avian transforming retrovirus isolated from a nephroblastoma that
RX carries the fos gene as the oncogene."
RL J. Virol. 61:3733-3740(1987).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M18041; AAA66158.1; ALT_INIT.
DR PIR: B34071; TVFV4.
DR HSSP: P01100; IFOS.
DR TRANSFAC: T01453; -
DR PROSITE: PS00036; BZIP_BASIC; 1.
DR PFAM: PF00170; bzip; 1.
KW Transforming protein; Nuclear protein; Phosphorylation; DNA-binding.
FT DNA_BIND 93
FT DOMAIN 119
FT DOMAIN 147
FT DOMAIN 147
SQ SEQUENCE 322 AA; 34333 MW; 7889C9B1 CRC32;
FOS_AVINK Length: 322 February 14, 2000 08:02 Type: P Check: 9324

1 SODFCIDLAV SSANFVPTV AISTSDIOW LVQPTLISSV ABSONGHPY
51 GVPAPAPPPA YSRPAVLKAP GGGGSGIGR GKVEQSPPEE EKKRRIRER
101 NKMAAKCKN RRRLDITLQ AETDQLEER SAQOAEIANL LKEKELFEI
151 LAHRPACMA PEELRFEEL AAATALDGA PSPAAAEAV ALPLMTEAP
201 AVPEPEPGS GLEKAPFD EILFSGPRE ASRSVPMDL PGASSFYASD
251 WPLGASGQ ELPICTPYV TCTPCSTYT STVFYTPFA DAFPCAAMH
301 RKGSSNPPS SDSISFTLL AL
11AA_SEQUENCE 1.0

```

ID FOS_CHICK STANDARD: PRT: 367 AA.
AC P11939;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN.
GN FOS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88262231.
RA MOEDERS H., JENWEIN T., ADAMKIEWICZ J., MUELLER R.;
RT "Isolation and structural analysis of a biologically active chicken
RT c-fos cDNA: identification of evolutionarily conserved domains in
RT fos protein.";
RL Oncogene 1:377-385(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88062857.
RA FUJIMURA K.I., ASHIDA K., NISHINA H., IBA H., MIYAJIMA N.,
RA NISHIZAWA M., KAWAI S.;
RT "The chicken c-fos gene: cloning and nucleotide sequence analysis.";
RL J. Virol. 61:4012-4018(1987).
CC - FUNCTION: C-FOS IS A NUCLEAR PHOSPHOPROTEIN WHICH FORM A TIGHT BUT
CC NON-COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR.
CC - SUBCELLULAR LOCATION: NUCLEAR.
CC - INDUCTION: C-FOS EXPRESSION INCREASES RAPIDLY UPON GROWTH FACTOR
CC STIMULATION OR WOUNDING OF CULTURED CELLS.
CC - SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M37000; AAA48670.1; -
DR EMBL: M18043; AA476823.1; -
DR PIR: A28368; TYCHFS.
DR HSSP: P01100; 1FOS.
DR TRNSFAC: T00123; -
DR PROSITE: PS00036; BZIP_BASIC: 1.
DR FRAM: PF00170; bzip.1.
KM PROTO-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
FT DNAS_BIND 138 159 BASIC MOTIF.
FT DOMAIN 164 192 LEUCINE-ZIPPER.
FT CONFLICT 90 90 N -> D (IN REF. 2).
SQ SEQUENCE 367 AA; 39004 MW; A3FCDACB CRC32;
FOS_CHICK Length: 367 February 14, 2000 08:02 Type: P Check: 7007 ..

ID FOS_HUMAN STANDARD: PRT: 380 AA.
AC P01100;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (GOS7 PROTEIN).
GN FOS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 83221560.
RA VAN STRATEN F., MULLER R., CURRAN T., VAN BEVEREN C., VERMA I.M.;
RT "Complete nucleotide sequence of a human c-onc gene: deduced amino
RT acid sequence of the human c-fos protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3183-3187(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.05 ANGSTROMS) OF 139-198 IN COMPLEX WITH JUN.
RX MEDLINE: 95115802.
RA GLOVER J.N., HARRISON S.C.;
RT "Crystal structure of the heterodimeric bzip transcription factor
RT c-fos-c-jun bound to DNA.";
RL Nature 373:257-261(1995).
CC - FUNCTION: C-FOS IS A NUCLEAR PHOSPHOPROTEIN WHICH FORM A TIGHT BUT
CC NON-COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR.
CC - SUBCELLULAR LOCATION: NUCLEAR.
CC - INDUCTION: C-FOS EXPRESSION INCREASES RAPIDLY UPON GROWTH FACTOR
CC STIMULATION OR WOUNDING OF CULTURED CELLS.
CC - SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01512; CAA24756.1; -
DR EMBL: K06550; AAA52471.1; -
DR PIR: A01342; TVHDF1.
DR PDB: 1FOS; 1O-JUL-95.
DR TRNSFAC: T00123; -
DR MIM: 164810; -
DR PROSITE: PS00036; BZIP_BASIC: 1.
DR FRAM: PF00170; bzip.1.
KM PROTO-oncogene; Nuclear protein; Phosphorylation; DNA-binding;
FT DNAS_BIND 139 160 BASIC MOTIF.
FT DOMAIN 165 193 LEUCINE-ZIPPER.
FT CONFLICT 380 AA; 40695 MW; CDDAEF5C CRC32;
SQ SEQUENCE 380 AA; 40695 MW; CDDAEF5C CRC32;
FOS_HUMAN Length: 380 February 14, 2000 08:02 Type: P Check: 4677 ..

11AA_SEQUENCE 1.0

11AA_SEQUENCE 1.0

ID FTRC_MAIZE STANDARD; PRT; 152 AA.
 AC P41347;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE FERREDOXIN-THIOREDOXIN REDUCTASE, CATALYTIC CHAIN PRECURSOR
 DE (RC 1.18.-.-) (FTR-C) (FERREDOXIN-THIOREDOXIN REDUCTASE SUBUNIT B)
 DE (FTR-B).
 OS FTRC.
 GN Zea mays (Maize).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Zea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94002243.
 RA MARG-MARTIN S., SPIELMANN A., STUTZ E., SCHERMAN P.;
 RT "Cloning and sequencing of a corn (Zea mays) nuclear gene coding for
 RT the chloroplast specific catalytic subunit of ferredoxin-thioredoxin
 RT reductase".
 RT Biochim Biophys. Acta 1183:207-209(1993).
 CC -1- FUNCTION: FTR IS A (4FE-4S) PROTEIN PLAYING A CENTRAL ROLE IN THE
 CC FERREDOXIN/THIOREDOXIN REGULATORY CHAIN. IT CONVERTS AN ELECTRON
 CC SIGNAL (PHOTOREDOXED FERREDOXIN) TO A THIOL SIGNAL (REDUCED
 CC THIOREDOXIN) IN THE REGULATION OF ENZYMES BY REDUCTION OF SPECIFIC
 CC DISULFIDE GROUPS. CATALYZES THE LIGHT-DEPENDENT ACTIVATION OF
 CC SEVERAL PHOTOSYNTHETIC ENZYMES.
 CC -1- SUBUNIT: HETERODIMER OF SUBUNIT A (VARIABLE SUBUNIT) AND SUBUNIT
 CC B (CATALYTIC SUBUNIT).
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X73549; ; NOT_ANNOTATED_CDS.
 DR MAIZEDB: 61547; ;
 KW Oxidoreductase; Iron-sulfur; 4Fe-4S; Chloroplast; Transit peptide.
 FT TRANSIT 1 38 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 1 38 FERREDOXIN-THIOREDOXIN REDUCTASE.
 FT METAL 91 152 CATALYTIC CHAIN.
 FT METAL 91 123 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 93 110 REDOX-ACTIVE (BY SIMILARITY).
 FT METAL 110 110 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 112 112 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 121 121 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SO SEQUENCE 152 AA; 16740 MW; 25FEE37 CRC32;
 FTRC_MAIZE Length: 152 February 14, 2000 08:02 Type: P Check: 4642 ..
 1 MSTVTTTGG CGGLPVRPLS TATRGPRRC AVRAQAAGAD ASNDKSVEYM
 51 RRFSEYARR SNTFFCADKT VTAIVYKGLA DHRDILGAPL CPCRHDDKA
 101 AEVAQGFVNC PCVPMREKRE CHOMLELPD NDFAKXQOVI SFEIKKANS
 151 KF
 11AA_SEQUENCE 1.0
 ID GDF8_BOVIN STANDARD; PRT; 375 AA.
 AC O18836; 018829;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 39, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 DE GDF8 OR MSTN OR MH.
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

OC Bovinae; Bos..
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANT MH TYR-313.
 RC STRAIN-FRIESIAN; TISSUE=MUSCLE, AND EMBRIO;
 RX MEDLINE: 97458167
 RA KAMBAJUR R., SHARMA M., SMITH T.P.L., BASS J.J.;
 RT "Mutations in myostatin (GDF8) in double-muscled Belgian Blue and
 RT Piedmontese cattle".
 RT Genome Res. 7:910-916(1997).
 RN [2]
 RP SEQUENCE FROM N.A. AND VARIANTS MH LEU-94 AND TYR-313.
 RC STRAIN-HOLSTEIN; TISSUE=SKLETAL MUSCLE;
 RX MEDLINE: 98024153.
 RA MCPHERON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene".;
 RT Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: WIDELY EXPRESSED THROUGHOUT DEVELOPMENT.
 CC LOW LEVELS ARE FOUND UP TO DAY 29 EMBRYOS. LEVELS INCREASE FROM
 CC DAY 31 UP UNTIL LATE GESTATION.
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN DEVELOPING AND ADULT
 CC SKELETAL MUSCLE. HIGHEST LEVELS FOUND IN THE HINDLIMB MUSCLES
 CC M. SEMIMEMBRANOSUS AND M. BICEPS FEMORIS; LOW LEVELS IN OTHER
 CC HINDLIMB MUSCLES.
 CC -1- DISEASE: DEFECTS IN GDF8 ARE THE CAUSE OF THE DOUBLE-MUSCLE
 CC PHENOTYPE OR MUSCULAR HYPERTROPHY (MH), AN AUTOSOMAL RECESSIVE
 CC DISEASE FREQUENTLY FOUND IN THE BELGIAN BLUE AND PIEDMONTSE
 CC CATTLE BREDS. THIS DISPAR IS CHARACTERIZED BY AN INCREASED
 CC NUMBER OF MUSCLE FIBERS (HYPERPLASIA), RESULTING IN AN INCREASE IN
 CC MUSCLE MASS OF 20-25%
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF019761; AAB81508.1; ;
 DR EMBL: AF019620; AAB86687.1; ;
 DR HSSP: P18075; IBMF.
 DR PROSITE: PS00250; TGF-BETA; 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal; Disease mutation.
 FT SIGNAL 1 266
 FT PROPEP 1 266
 FT CHAIN 1 375
 FT DISULFID 267 375 GROWTH/DIFFERENTIATION FACTOR 8.
 FT DISULFID 281 340 BY SIMILARITY.
 FT DISULFID 309 372 BY SIMILARITY.
 FT DISULFID 313 374 BY SIMILARITY.
 FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 47 47
 FT CARBOHYD 71 71
 FT VARIANT 94 94
 FT VARIANT 94 94
 FT VARIANT 313 313 F > L (IN MH; PIEDMONTSE BREED).
 FT VARIANT 14 14 C > Y (IN MH; PIEDMONTSE BREED).
 FT CONFLICT 14 14 T > M (IN REF. 2).
 SO SEQUENCE 375 AA; 42520 MW; F97E649A CRC32;
 GDF8_BOVIN Length: 375 February 14, 2000 08:02 Type: P Check: 9403 ..
 1 MOKIQTIVY YLFTLLVAGP VDLNSENSEK ENVEKEGICN ACIWRNNTS
 51 SRLIAKIQDI LSKRLBTAP NISKDAIROL LPKAPPLLEL IDQFDORDA
 101 SSDGSLDDD YHARTETVIT MPTESDLILQ VEGAPKCCFE KFSKSIQYVK
 151 LYKAQIMWYL RPYKTPATVE VOILRLIKPM KDGRTYRGIR SKLDMNGGT
 201 GIMQSIDVKT VLOWMLKQPE SMIGIEIKAL DENGHDLAVT FPEPEDDGLT

```

251 PELEVKYDT PKRSRDFGL DCDHSTSR CCRRPLTVDF EAFGDMWII
301 PRRYKANCY GECEFEVLQK YPHTHLVHQ NPGSAGPCC TPTKSPIM
351 LYFNGEGOI YKIPAMVVD RCGCS

!!AA_SEQUENCE 1.0
ID GDF8_HUMAN STANDARD: PRT; 375 AA.

AC 014793:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 98024153.
RA MCPHERRON A.C., LEE S.-J.;
RT "Double muscling in cattle due to mutations in the myostatin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF019627; AAB8694.1; -
DR MIM; 601788; -
DR PROSITE; PS00250; TGF-BETA; 1.
DR PFM; PF00019; TGF-beta; 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 266
FT PROPEP 2 375
FT CHAIN 267 375
FT DISULFID 281 340
FT DISULFID 309 372
FT DISULFID 313 374
FT DISULFID 339 339
FT CARBOHYD 71 71
SQ SEQUENCE 375 AA; 42750 MW; D0EC1323 CRC32;

GDF8_HUMAN Length: 375 February 14, 2000 08:02 Type: P Check: 1814

```

```

AC 008689:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 97284412.
RA MCPHERRON A.C., LAWLER A.M., LEE S.-J.;
RT "Regulation of skeletal muscle mass in mice by a new TGF-beta
RT superfamily member.";
RL Nature 387:83-90(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN DEVELOPING AND ADULT
CC SKELETAL MUSCLE. WEAK EXPRESSION IN ADIPOSE TISSUE.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT DAY 9.5 POST-CONITUM IN
CC ONE-THIRD OF DEVELOPING SOMITES. AT DAY 10.5, EXPRESSED IN THE
CC MYOTOME COMPARTMENT OF SOMITES. AT LATER STAGES OF DEVELOPMENT,
CC DETECTED IN A WIDE RANGE OF DEVELOPING MUSCLES. EXPRESSION
CC CONTINUES IN ADULTHOOD.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U84005; AAC3167.1; -
DR HSSP; P18075; BMP.
DR MGD; MGI:95691; MSTN.
DR PROSITE; PS00250; TGF-BETA; 1.
DR PFM; PF00019; TGF-beta; 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 267
FT PROPEP 2 376
FT CHAIN 268 341
FT DISULFID 282 341
FT DISULFID 310 373
FT DISULFID 314 375
FT DISULFID 340 340
FT CARBOHYD 72 72
SQ SEQUENCE 376 AA; 42921 MW; F1ED196 CRC32;

GDF8_MOUSE Length: 376 February 14, 2000 08:02 Type: P Check: 2293

```

```

!!AA_SEQUENCE 1.0
ID GDF8_MOUSE STANDARD: PRT; 376 AA.

```

```

1 MOKLQCVYI YLFMLIVAGP VDLNENSEK ENVEKEGLCN ACTWRONTKS
51 SRIAIKIQI LSKRLLETAP NISKDVIRQL LPRAPPLREL IDQYDVORD
101 DSDSLDDDD YHATETIIT MPTESDFLM VDGKPKCFE KFSKIQYNK
151 VYKQQLMIYI RPEVPTIVF VOILRLIKP KDGTRTGAR SLKLDMMNGI
201 GIMOSIDVKT VLQNLKQPE SNLGEIKAL DENGHDNAV FPGGEGDLN
251 PFLEVKYDTI PKRSRDFGL DCDHSTSR CCRRPLTVDF EAFGDMWII
301 PRRYKANCY GECEFEVLQK YPHTHLVHQ NPGSAGPCC TPTKSPIM
351 LYFNGEGOI YKIPAMVVD RCGCS

```

```

!!AA_SEQUENCE 1.0
ID GDF8_PAPHA STANDARD: PRT; 375 AA.

```

```

1 MOKLQCVYI YLFMLIAGP VDLNENSEK ENVEKEGLCN ACTWRONTKS
51 YSRIAIKIQI LSKRLLETAP NISKDVIRQL LPRAPPLREL IDQYDVORD
101 DSDSLDDDD DYHATETIIT MPTESDFLM QADGKPKCFE KFSKIQYN
151 KYVKAQLMIYI LRPVKTPTIV VOILRLIKP KDGTRTGAR SLKLDMMSPG
201 TGIMOSIDVKT VLQNLKQPE ESNLGEIEKA LDENGHDNAV TFGGEGEGL
251 NPFLEVKYDTI PKRSRDFGL DCDHSTSR CCRRPLTVDF EAFGDMWII
301 APRRYKANCY GECEFEVLQK YPHTHLVHQ NPGSAGPCC TPTKSPIN
351 MLYFNGEGOI IYKIPAMVVD DRGCS

```

```

AC 018928;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN.
OS Papilio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Papio.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 98024153.
RA MCPHERSON A.C., LEE S.-I.;
RT "Double muscling in cattle due to mutations in the myostatin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF019619; AAB86686.1;
CC DR PROSITE: PS00250; TGF-BETA; 1.
CC DR PFAM: PF00019; TGF-Beta; 1.
CC KW Growth factor; Cytokine; Glycoprotein; Signal.
CC FT SIGNAL 1 ?
CC FT PROPEP 267 266 POTENTIAL.
CC FT CHAIN 267 375 GROWTH/DIFFERENTIATION FACTOR 8.
CC FT DISULFID 281 340 BY SIMILARITY.
CC FT DISULFID 309 372 BY SIMILARITY.
CC FT DISULFID 313 374 BY SIMILARITY.
CC FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 71 71 POTENTIAL.
CC SQ SEQUENCE 375 AA; 42688 MW; 61C8D2B7 CRC32;

GDF8_PAPHA Length: 375 February 14, 2000 08:02 Type: P Check: 1463 ..

1 MOKLQIYVI YLPMLIYAGP VDLNENSEQK ENVEKEGLCN ACTWRQNTKS
51 SRIEAIKIQI LSKRLLETAP NISKDAIRQL LPRAPPLREL IDQYDVQRD
101 SSDGSLDDDD YHATTEITIT MPTESDLMO VDGKPKCFE KFSKRIQYNK
151 YKQAQLWLYL RPYKPTIYF VOILRLIKPM KDGRTYTGIR SLKIDMNGT
201 GIMOSIDVKT VLQNLKQPE SNLGEIKAL DENGHDLAVT FPGEGEDGLN
251 PLEVKYVTDI PKRSRDFGL DCDHSTESR CCRYPITYDF EALGWDWIIA
301 PKRYKANVCS GECEYFVLQK YPHTHLVHOA NPGSAGPCG TPTKMSPINM
351 LYFNGKEQII YGKIPAMVVD RCGCS

11AA_SEQUENCE 1.0
ID GDF8_PIG STANDARD: PRT: 375 AA.
AC 018931;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
RC TISSUE-SKELETAL MUSCLE;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 98024153.
RA MCPHERSON A.C., LEE S.-I.;
RT "Double muscling in cattle due to mutations in the myostatin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF019623; AAB86690.1;
CC DR PROSITE: PS00250; TGF-BETA; 1.
CC DR EMBL: AF033855; AAC08035.1;
CC DR EMBL: AF033798; AAC02489.1;
CC DR PROSITE: PS00250; TGF-BETA; 1.
CC DR PFAM: PF00019; TGF-Beta; 1.
CC KW Growth factor; Cytokine; Glycoprotein; Signal.
CC FT SIGNAL 1 ?
CC FT PROPEP 267 266 POTENTIAL.
CC FT CHAIN 267 375 GROWTH/DIFFERENTIATION FACTOR 8.
CC FT DISULFID 281 340 BY SIMILARITY.
CC FT DISULFID 309 372 BY SIMILARITY.
CC FT DISULFID 313 374 BY SIMILARITY.
CC FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 71 71 POTENTIAL.
CC SQ SEQUENCE 375 AA; 42791 MW; 65CB82B8 CRC32;

GDF8_PIG Length: 375 February 14, 2000 08:02 Type: P Check: 1805 ..

1 MOKLQIYVI YLPMLIYAGP VDLNENSEQK ENVEKEGLCN ACMMRQNTKS
51 SRIEAIKIQI LSKRLLETAP NISKDAIRQL LPRAPPLREL IDQYDVQRD
101 SSDGSLDDDD YHATTEITIT MPTESDLMO VEGKPKCFE KFSKRIQYNK
151 YKQAQLWLYL RPYKPTIYF VOILRLIKPM KDGRTYTGIR SLKIDMNGT
201 GIMOSIDVKT VLQNLKQPE SNLGEIKAL DENGHDLAVT FPGEGEDGLN
251 PLEVKYVTDI PKRSRDFGL DCDHSTESR CCRYPITYDF EAFGWDWIIA
301 PKRYKANVCS GECEYFVLQK YPHTHLVHOA NPGSAGPCG TPTKMSPINM
351 LYFNGKEQII YGKIPAMVVD RCGCS

11AA_SEQUENCE 1.0
ID GDF8_RAT STANDARD: PRT: 376 AA.
AC 035312;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
RC TISSUE-SKELETAL MUSCLE;

```

```

RX MEDLINE: 98024153;
RA MCPHERSON A.C., LEE S.-J.;
RT "Double muscling in cattle due to mutations in the myostatin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997);
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF019624; AAB86691.1;
DR PROSITE: PS00250; TGF-BETA: 1.
DR PFAM: PF00019; TGF-beta: 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 267
FT PROPEP 1 267
FT CHAIN 268 376
FT DISULFID 282 341
FT DISULFID 310 373
FT DISULFID 314 375
FT DISULFID 340 340
FT CARBOHYD 72 72
SQ SEQUENCE 376 AA; 42829 MW; BFC16D96 CRC32;

GDF8_RAT Length: 376 February 14, 2000 08:02 Type: P Check: 1496
1 MQRKPMVYV IYLFVLIAG PVDLNESER EANVEKESLC MACAMRONR
51 YSRLEAIKIQ ILSKRLLETA PNISKDAIRO LLPRAPLURE LIDQYDVRD
101 SSSDGSLEDD DYHATTEIIL TMPTESDFLM QADGPKKCF KFKSKIQYN
151 KYVKAQLMIY LRAVKTPTTV FVQILRLIKP MKDGRYTG I RSLKLDMSFG
201 TGWGSIDVK TVLQNLKOP ESNLGEIKA LDENGHDVAV TFGPGEGDL
251 NFPLEKVTID TPKRSRDFG LDCDEHSTES ROCRPLPLVD FEAFGDWII
301 APRRYANVC SCEPEVFLQ KYPHHLVHQ ANPRGSAGPC CPTKMSPIN
351 LYFNCKEOL IYKIPAMVY DRGCS

11AA_SEQUENCE 1.0
ID GDF8_SHEEP STANDARD: PRT: 375 AA.
AC 018830;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 98024153;
RA MCPHERSON A.C., LEE S.-J.;
RT "Double muscling in cattle due to mutations in the myostatin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997);
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF019624; AAB86691.1;
DR PROSITE: PS00250; TGF-BETA: 1.
DR PFAM: PF00019; TGF-beta: 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 266
FT PROPEP 1 266
FT CHAIN 267 375
FT DISULFID 281 340
FT DISULFID 309 372
FT DISULFID 313 374
FT DISULFID 339 339
FT CARBOHYD 48 48
FT CARBOHYD 71 71
SQ SEQUENCE 375 AA; 42827 MW; B36084EE CRC32;

GDF8_SHEEP Length: 375 February 14, 2000 08:02 Type: P Check: 1548
1 MQRKQIVYI YLFMLVAGP VDLNENSEOK ENVEKKGLCN ACLMRONKS
51 SRLEAIKIQ ILSKRLLETA PNISKDAIRO LLPRAPLURE LIDQYDVRD
101 SSSDGSLEDD YHATTEIIV TMPTESDLAE VQEKPKKCF KFKSKIQHNK
151 VYKQQLMIY LRAVKTPTTV FVQILRLIKP MKDGRYTG I RSLKLDMSFG
201 GIMOSIDVK TVLQNLKOP ESNLGEIKA LDENGHDVAV TFGPGEGDL
251 NFPLEKVTID TPKRSRDFG LDCDEHSTES ROCRPLPLVD FEAFGDWII
301 PKRYANVC SCEPEVFLQ KYPHHLVHQ ANPRGSAGPC CPTKMSPIN
351 LYFNCKEOL IYKIPAMVY DRGCS

11AA_SEQUENCE 1.0
ID GDF8_PANTR STANDARD: PRT: 123 AA.
AC 028914;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLYCOPHORIN B PRECURSOR.
GN GYBP OR GPB.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96029146.
RA HUANG C.H., XIE S.S., SOCHA W., BLUMENFELD O.O.;
RT "Sequence diversification and exon inactivation in the glycophorin A
RT gene family from chimpanzee to human.";
RL J. Mol. Evol. 41:478-486(1995).
CC -1- FUNCTION: THIS PROTEIN IS A MINOR SIALOGLYCOPROTEIN IN ERYTHROCYTE
CC MEMBRANES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE GLYCOPHORIN A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S79726; AAB35339.1;

```

DR HSP: P02724; 1MSR.
 DR PROSITE; P500312; GLYCOPHORIN A; 1.
 DR PFAM; PF01102; Glycophorin A; 1.
 KW Erythrocyte; Transmembrane; Sialic acid; Glycoprotein; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 123 GLYCOPHORIN B.
 FT TRANSMEM 93 113 POTENTIAL.
 SO SEQUENCE 123 AA; 13246 MW; 28850255 CRC32.

GPB_PANTR Length: 123 February 14, 2000 08:02 Type: P Check: 149

1 MGKRIIFUL LSEIVSISAS STEEVAMHS TSSSVTSYI SSOTNDKMG
 51 DTPATLGAH EVSEISVTV YPEEDNGEM VQVHPSPRP APVYIILIL
 101 CMAAGVIGTI LLISYIGRL IKA

!!AA_SEQUENCE 1.0 STANDARD; PRT; 339 AA.
 ID GPB_ECOLI
 AC P37606;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLYCEROL-3-PHOSPHATE DEHYDROGENASE [NAD+] (EC 1.1.1.8).
 GN GPAS.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

!!
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA MEDLINE; 94316500.
 RX SORIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes."
 RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -1- CATALYTIC ACTIVITY: NAD-GLYCEROL-3-PHOSPHATE + NAD(+) -
 CC GLYCERONE PHOSPHATE + NADH
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE)
 CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U00039; CAB34639.1;
 DR EMBL; AE000439; AAC76632.1;
 DR ECGENE; EG20091; GPAS.
 DR PROSITE; PS00957; NAD_G3PDH; 1.
 DR PFAM; PF01210; NAD_GLY3P_dh; 1.
 KW Oxidoreductase; NAD.
 SO SEQUENCE 339 AA; 36361 MW; 085E0E12 CRC32;

GPB_ECOLI Length: 339 February 14, 2000 08:02 Type: P Check: 4443

1 MNQNASMTV IGAGSYGTAL AITLARGHE VLMGDPPEH TATLDRDGN
 51 AAFIPDPFP DTLHESDIA TALAASRNIL VVPSHVGGE VLRQIKPLMR
 101 PDAFLVATK GLEAETGRLL QDVAREALGD QPLAVISGG TPAKELANGL
 151 PTAISLASTD QTFADLQQL LHGKSEFRV SNPDFGYOL GGAANKVIAI
 201 GAGNSDIGF GANARTALIT KGLAEMRLG ALGADPAF WGMAGLDLV
 251 LCTDNOSRN RREGMLGOG MDVOSAOKI GQVVEGYRNT KEVELAHRE
 301 GVEMPTTEEI YQVLYCGKNA REALLTILGR ARKDESSH

!!AA_SEQUENCE 1.0 STANDARD; PRT; 355 AA.
 ID GPRD_HUMAN
 AC P49238;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE G PROTEIN-COUPLED RECEPTOR GPR13 (V28) (BETA CHEMOKINE
 DE RECEPTOR-LIKE 1) (CMK-BRL-1) (CMKBLR1).
 GN GPR13.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96011651.
 RA RAPPORT C.J., SCHWEICKART V.L., EDDY R.L. JR., SHOWS T.B., GRAY P.W.;
 RT "The orphan G-protein-coupled receptor-encoding gene v28 is closely
 RT related to genes for chemokine receptors and is expressed in lymphoid
 RT and neural tissues."
 RL Gene 163:295-299(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95374679.
 RA COMBADIÈRE C., AHUJA S.K., MURPHY P.M.;
 RT "Cloning, chromosomal localization, and RNA expression of a human
 RT beta chemokine receptor-like gene."
 RL DNA Cell Biol. 14:673-680(1995).
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND NEURAL TISSUES.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U20350; AAA91783.1;
 DR EMBL; U28934; AAA87032.1;
 DR GCRDB; GCR_1992;
 DR GIM; 601470;
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 1.
 KW G-protein coupled receptor; Transmembrane.

FT DOMAIN 1 31 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 32 59 1 (POTENTIAL).
 FT DOMAIN 60 69 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 70 90 2 (POTENTIAL).
 FT DOMAIN 91 103 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 104 125 3 (POTENTIAL).
 FT DOMAIN 126 142 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 143 167 4 (POTENTIAL).
 FT DOMAIN 168 195 5 (POTENTIAL).
 FT TRANSMEM 196 215 6 (POTENTIAL).
 FT DOMAIN 216 231 7 (POTENTIAL).
 FT TRANSMEM 232 256 8 (POTENTIAL).
 FT DOMAIN 257 273 9 (POTENTIAL).
 FT TRANSMEM 274 297 10 (POTENTIAL).
 FT DOMAIN 298 355 11 (POTENTIAL).
 FT DISULFID 102 175 BY SIMILARITY.
 SO SEQUENCE 355 AA; 40396 MW; E809E31D CRC32;

GPRD_HUMAN Length: 355 February 14, 2000 08:02 Type: P Check: 2863

1 MDQFESVTE NFEYDLAEA CYIGIIVFG TVFISIVSY IFAIGLVGL
 51 LVVFLATNSK KPKSVTDIYL LNLALSDLE VALPFWTHY LNEGGLHNA

101 MCKETTAFF IGFSGISFFI TVISIDRYLA IYLAANSNN RTVOHGVTIS
 151 LGWMAAILV APOPMFTKQ KENECLGDP EYLOEIMFVL RNVEINFLG
 201 LPLLLMSYC YRIIQTLES CKNRKAKAI KLILVIVF FLEWTPYNYM
 251 IFETIKLYD FEPSCDMRKD LRLALSYTE VAFSHCCINP LIYAFAGEKF
 301 RRYLYHLYGK CLAVLCGRSV HYDFSSSESO RSRHGSVLSS NFIYHSDGD
 351 ALLLL

11AA-SEQUENCE 1.0 STANDARD: PRT: 354 AA.
 ID GPRD_RAT
 AC P35411
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE G-PROTEIN-COUPLED RECEPTOR GPR13 (RBS11).
 GN GPR13.
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-SPRAGUE-DAWLEY; TISSUE-SPINAL CORD;
 RX MEDLINE: 94323113.
 RA HARRISON J.R., BARBER C.M., LYNCH K.R.;
 RT "cDNA cloning of a G-protein-coupled receptor expressed in rat spinal
 cord and brain related to chemokine receptors.";
 RL Neurosci. Lett. 169:85-89(1994).
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN ADULT SPINAL CORD, BRAIN,
 CC KIDNEY, GUT, UTERUS AND TESTES.
 CC -1- PFM: THIS PROTEIN IS NOT N-GLYCOSYLATED WHICH IS UNUSUAL FOR
 CC G-PROTEIN-COUPLED RECEPTORS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U04808; AAB87093.1; -.
 DR GCRDB: GCR_0268; -.
 DR PROSITE: PS00237; G-PROTEIN-RECEPTOR; 1.
 DR PFM: PFM0001; 7tm_1; 1.
 KW G-protein coupled receptor; Transmembrane.
 FT DOMAIN 1 32
 FT TRANSMEM 33 60
 FT DOMAIN 61 70
 FT TRANSMEM 71 91
 FT DOMAIN 92 104
 FT TRANSMEM 105 126
 FT DOMAIN 127 143
 FT TRANSMEM 144 168
 FT DOMAIN 169 196
 FT TRANSMEM 197 216
 FT DOMAIN 217 232
 FT TRANSMEM 233 257
 FT DOMAIN 258 274
 FT TRANSMEM 275 298
 FT DOMAIN 299 354
 FT TRANSMEM 354 403
 FT DISULFID 103 176
 FT SEQUENCE 354 AA; 40327 MW; 798DA277 CRC32;
 BY SIMILARITY

GPRD_RAT Length: 354 February 14, 2000 08:02 Type: P Check: 4478
 1 MPTSPEDLD ENFEYDSAE ACYLGDIYAF GTIFLSIFYS LYTFELGVGN

51 LLVVALTNS RKSRSITDIY LLNLALSDDL EVATLPFWTH YLISHESLHN
 101 AMCKLTAF IGFEGGIFE ITVISIDRYL AYLANSNN RTVOHGVTI
 151 SLGWMAAIL VASPOPMFTK RKNCECLGDP EYLOEIMFVL RNSEVNLG
 201 FVPLLLMSF CYFRIVRTLF CKNRKAKA IRILIVVYV FLEWTPYNI
 251 VIFETIKFY NFPPSCGMKR DLRLALSYTE VAFSHCCINP PIYAFAGEK
 301 FRYLRHLIN KCLAVLCGRP VHAQSTIESQ RSRHDSILSS LHYTSGEG
 351 SLLL

11AA-SEQUENCE 1.0 STANDARD: PRT: 247 AA.
 ID GRAB_MOUSE
 AC P04187;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GRANZYME B(G,H) PRECURSOR (EC 3.4.21.79) (CYTOTOXIC CELL PROTEASE 1)
 DE (CCP1) (CTLA-1) (FRAGMENTIN 2).
 GN GZMB OR CTLA OR CTLA-1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 86208120.
 RA LOBE C.G., FINLAY B.B., PARANCHYCH M., PATRAU V.H., BLECKLEY R.C.;
 RT "Novel serine proteases encoded by two cytotoxic T
 RT lymphocyte-specific genes.";
 RL Science 252:858-861(1996).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 89062424.
 RA LOBE C.G., UPTON C., DUGGAN B., EHMAN N., LETELLIER M., BELL J.,
 RA MCFADDEN G., BLECKLEY R.C.;
 RT "Organization of two genes encoding cytotoxic T lymphocyte-specific
 RT serine proteases CCP1 and CCP1.";
 RL Biochemistry 27:6941-6946(1988).
 [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 86284960.
 RA BRUNET J.F., DOSSETO M., DENIZOT F., NATTEI M.-G., CLARK W.R.,
 RA HAOI T.M., FERRIER P., NABHOLZ M., SCHMITT-VERHULST A.M.,
 RA LUCIANI M.F., GOLSTEIN P.;
 RT "The inducible cytotoxic T-lymphocyte-associated gene transcript
 RT CTLA-1 sequence and gene localization to mouse chromosome 14.";
 RL Nature 322:268-271(1986).
 [4]
 RN SEQUENCE OF 227-247 FROM N.A.
 RP STRAIN-C57BL/6J;
 RX MEDLINE: 94319082.
 RA KO W.S., WANG X., HORTON J.H., HAGEN M.D., TAKAHASHI N.,
 RA VAEZAKI Y., NADDEVU J.H.;
 RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";
 RL Mamm. Genome 5:349-353(1994).
 [5]
 RN SEQUENCE OF 21-40.
 RX MEDLINE: 87215932.
 RA MASSON D., TSCHOP J.;
 RT "A family of serine esterases in lytic granules of cytolytic T
 RT lymphocytes.";
 RL Cell 49:679-685(1987).
 [6]
 RN 3D-STRUCTURE MODELING.
 RP MEDLINE: 89184501.
 RA MORPHY M.E.P., MOULT J.J., BLECKLEY R.C., GERSHENFELD H.,
 RA WEISSMAN I.L., JAMES M.N.G.;
 RT "Comparative molecular model building of two serine proteinases from
 RT cytotoxic T lymphocytes.";

RL Proteins 4:190-204(1988).
 CC -1- FUNCTION: THIS ENZYME IS NECESSARY FOR TARGET CELL LYSIS IN CELL-MEDIED IMMUNE RESPONSES. IT CLEAVES AFTER ASP. SEEMS TO BE LINKED TO AN ACTIVATION CASCADE OF CASPASES (ASPARTATE-SPECIFIC CYSTEINE PROTEASES) RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES CASPASE-3, -7, -9 AND 10 TO GIVE RISE TO ACTIVE ENZYMES MEDIATING APOPTOSIS (BY SIMILARITY).
 CC CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-|-XAA >> ASP-|-XAA > MET-|-XAA, SER-|-XAA.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC T-LYMPHOCYTES AND NATURAL KILLER CELLS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMS AND TO MASI CELL PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X04072; CAA27715.1; -
 DR EMBL: M12302; AAA37383.1; -
 DR EMBL: M22526; AAB61756.1; -
 DR EMBL: U05707; AAB60470.1; -
 DR PIR: A00956; PRMSCL.
 DR PIR: A28952; A28952.
 DR PIR: B26944; B26944.
 DR PDB: 2CP1; 15-OCT-94.
 DR MGD: MGI:109267; GZMB.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PFAM: PF00089; trypsin; 1.
 DR KW Hydrolyase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
 KM Apoptosis; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 20 ACTIVATION PEPTIDE.
 FT CHAIN 21 247 GRANZYME B(G,H).
 FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 49 63 BY SIMILARITY.
 FT DISULFID 142 209 BY SIMILARITY.
 FT DISULFID 173 188 BY SIMILARITY.
 FT CARBOHYD 71 71 POTENTIAL.
 FT CARBOHYD 182 182 POTENTIAL.
 SQ SEQUENCE 247 AA; 27470 MW; E05916CB CRC32;
 GRAB_MOUSE Length: 247 February 14, 2000 08:02 Type: P Check: 92 ..
 1 MNLLLLLLT SLASRTAGE IIGGHEVKPH SRPYMALSI KDOQPEALICG
 51 GFLREDEVYL TAAHCEGSI NVTLGHNK EOKTQVIF MYKCIHPDY
 101 NKRTSNDIM LKLSKAKR TRAVRLNP RRVNVAPGD VCVVAGWGR
 151 APMGYSNLT QEVELYQKD RECESYFKNR YNKINQICAG DPRTKASR
 201 GSGGPLVCK KVAAGISYSG YKDSPPRAF TKVSSFLSWI KTKMSS
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 267 AA.
 ID HDHA_CLOSO
 AC P50200;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (BILE ACID 7-DEHYDROXYLASE) (7-ALPHA-HSDH).
 DE Clostridium sordeilii.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.

RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-18.
 RC STRAIN-ATCC 9714.
 RX MEDLINE: 94327451
 RA COLEMAN J.P., HUDSON L.L., ADAMS M.J.;
 RT "Characterization and regulation of the NADP-linked 7 alpha-hydroxysteroid dehydrogenase gene from Clostridium sordeilii.";
 RL J. Bacteriol. 166:4863-4874(1994).
 CC -1- CATALYTIC ACTIVITY: 3-ALPHA,7-ALPHA,12-ALPHA-TRIHYDROXY-5-BETA-CHOLANATE + NADP(+) = 3-ALPHA,12-ALPHA-DIHYDROXY-7-OXO-5-BETA-CHOLANATE + NADPH.
 CC -1- SUBUNIT: HOMOTETRAMER (PROBABLE).
 CC -1- INDUCTION: BY BILE ACID. EXPRESSED DURING GROWTH-PHASE.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L12058; AAA3556.1; -
 DR HSSP: P25529; 1AHH.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF00678; adh_short; 2.
 DR KW Oxidoreductase; NADP; Bile acid catabolism.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 158 158 BY SIMILARITY.
 SQ SEQUENCE 267 AA; 29177 MW; FCCD9D5A CRC32;
 HDHA_CLOSO Length: 267 February 14, 2000 08:02 Type: P Check: 1047 ..
 1 MNKLKVAL VTSATRGIGL ASAIKLAONG AIVYGVRL EATQELCKDY
 51 KEELGLIKPV FEDAYNIDY KEMIDITLKN ESKIDILVNN FGTRPEKDL
 101 DLYNGDEDIF FELFYNVGS VYRLSKLIP HMIENKGGSI VNISVSGSI
 151 PDISRIIGYGV SKSGVNNITK QIAIQYAKYG IKNVNLPLG IATDAMNSM
 201 PDEFKSFILS HVLNRIQNP EDIANSLVLF VPSEDSYIT GSILEVSGGY
 251 NLGTPQYAEF VQSKVYE
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 296 AA.
 ID HG2A_HUMAN
 AC P04233; Q29832;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 05-DEC-1999 (Rel. 39, Last annotation update)
 DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (HLA-DR ANTIGENS DE ASSOCIATED INVARLANT CHAIN) (P33) (CD74 ANTIGEN).
 GN CD74 OR DR4G.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86093681.
 RA KUDO J., CHAO L.Y., NARNI F., SAUNDERS G.F.;
 RT "Structure of the human gene encoding the invariant gamma-chain of class II histocompatibility antigens.";
 RL Nucleic Acids Res. 13:8827-8841(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84207945.
 RA STRUBIN M., MACH B., LONG E.O.;
 RT "The complete sequence of the mRNA for the HLA-DR-associated invariant chain reveals a polypeptide with an unusual transmembrane

RT POLARITY.";
 RL EMBL J. 3:869-872(1984).
 [3]
 RP SEQUENCE OF 27-96 FROM N.A.
 RX MEDLINE: 84170234.
 RA CLAESSON L., LARHAMMAR D., RASK L., PETERSON P.A.;
 RT "CDN clone for the human invariant gamma chain of class II
 RT histocompatibility antigens and its implications for the protein
 RT structure." Acad. Sci. U.S.A. 80:7395-7399(1983).
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7395-7399(1983).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86233451.
 RA O'SULLIVAN D.M., LARHAMMAR D., WILSON M.C., PETERSON P.A.,
 RA QUARANTA V.;
 RT "Structure of the human Ia-associated invariant (gamma)-chain gene:
 RT identification of 5' sequences shared with major histocompatibility
 RT complex class II genes." Proc. Natl. Acad. Sci. U.S.A. 83:4484-4488(1986).
 RL [5]
 RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS) OF 103-117.
 RX MEDLINE: 96085023.
 RA GHOSH P., AMAYA M., MELLINS E., WILEY D.C.;
 RT "The structure of an intermediate in class II MHC maturation: CLIP
 RT bound to HLA-DR3." Nature 378:457-462(1995).
 RL [6]
 RP STRUCTURE BY NMR OF 134-208.
 RX MEDLINE: 99059718.
 RA JASANOFF A., WAGNER G., WILEY D.C.;
 RT "Structure of a trimetric domain of the MHC class II-associated
 RT chaperonin and targeting protein II." EMBO J. 17:6812-6818(1998).
 RL [7]
 RP FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING
 CC BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A
 CC COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE
 CC COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE
 CC PEPTIDE LOADING OF CLASS II TAKES PLACE.
 CC [8]
 RP SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTRIMERS.
 CC [9]
 RP SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC [10]
 RP SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC [11]
 RP DATABASE: NAME-PROT: NOTE-CD guide CD74 entry;
 CC [12]
 RP WWW: http://www.ncbi.nlm.nih.gov/protomol/cd74.htm.
 CC [13]
 RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC [14]
 RP EMBL: X03339; CAA27046.1; -
 CC [15]
 RP EMBL: X03340; CAA27047.1; -
 CC [16]
 RP EMBL: X03144; AAA36304.1; -
 CC [17]
 RP EMBL: X00497; CAA25192.1; -
 CC [18]
 RP EMBL: X00497; CAA25193.1; -
 CC [19]
 RP PIR: A30060; HLHUG.
 DR PDB: 1A6A; 27-MAY-98.
 DR PDB: 1IE; 15-FEB-99.
 DR MIM: 142790; -
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 DR PRAM: PF00086; thyroglobulin_1; 1.
 KW Glycoprotein; MHC; Proteoglycan; Transmembrane; Alternative splicing;
 KW Signal-anchor; 3D-structure.
 FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 47 72 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 73 296 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 103 117 CLIP.
 FT CARBOHYD 130 130
 FT CARBOHYD 136 136
 FT CARBOHYD 282 282
 FT DOMAIN 228 271 THYROGLOBULIN TYPE I.

FT VARSPLIC 209 272 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 167 167 T -> R (IN REF. 2 AND 3).
 SO SEQUENCE 296 AA; 33460 MW; 4E8A044A CRC32;
 HG2A_HUMAN Length: 296 February 14, 2000 08:02 Type: P Check: 9949 ..
 1 MHRRSRSCR EDQKRYMDQ RDLSNNEDL PMLGRPGAP ESKSRGALY
 51 TGSILVTL LAGQATYAF LYQOQGRDLX LVTVSQNLQ ENLRMLPKP
 101 PKPYSKRMA TPLMALPM GALPGQMN AKKYNMED HYNHLLQAD
 151 PLKYPLKG SEPENLHLK NMTETLWKY FESNHHML FENSRHSLQ
 201 KPLDAPKVL TKQGEVSHI PAVHSGFRP KCDENGNIPL LQCGSLGIC
 251 WCFPNQTEV PNTSRGHN CSESLLEDP SSGLYTKOD LGPYPM
 11AA_SEQUENCE 1.0 STANDARD; PRT; 359 AA.
 ID HIS8_SALTY
 AC P10369;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-
 DE PHOSPHATE TRANSAMINASE).
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 89094829.
 RA CAROMONGO M.S., CHARIOTRI L., ALIFANO P., NAPPO A.G., BRUNI C.B.;
 RT "Structure and function of the Salmonella typhimurium and Escherichia
 RT coli K-12 histidine operons." J. Mol. Biol. 203:565-606(1988).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RA BARNES W.M., HUSON R.N., WHITTIER R.;
 RL Submitted (Aug-1989) to the EMBL/GenBank/DBJ databases.
 CC [3]
 RP CATALYTIC ACTIVITY: L-HISTIDINOL-PHOSPHATE + 2-OXOGLUTARATE =
 CC 3-(IMIDAZOL-4-YL)-2-OXOPROPYL PHOSPHATE + GLUTAMATE.
 CC [4]
 RP COFACTOR: PYRIDOXAL PHOSPHATE.
 CC [5]
 RP PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC [6]
 RP SUBUNIT: HOMODIMER.
 CC [7]
 RP SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC [8]
 RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC [9]
 RP EMBL: X13464; CAA31824.1; -
 CC [10]
 RP EMBL: J01804; AAA88616.1; -
 CC [11]
 RP PIR: J01584; XNEBHC.
 DR STUYGENE: SGI0159; HISG.
 DR PROSITE: PS00599; AA-TRANSFER CLASS_2; 1.
 DR PRAM: PF00222; aminotran_2; 1.
 KW Histidine biosynthesis; Transferase; Aminotransferase;
 KW Pyridoxal phosphate
 FT BINDING 217 217 PYRIDOXAL PHOSPHATE (PROBABLE).
 FT CONFLICT 148 164 DGKIVFVCSNNPTGQ -> TQKWCSEVAPILPBN
 FT CONFLICT 226 226 C -> R (IN REF. 2).
 FT CONFLICT 260 260 L -> S (IN REF. 1).
 FT CONFLICT 264 265 GI -> ES (IN REF. 1).

3D-STRUCTURE MODELING OF 136-196.
 MEDLINE: 94259190.
 RA NEIDLE S., GOODWIN G.H.;
 RT "A homology-based molecular model of the proline-rich homeodomain
 protein prh, from hematopoietic cells.";
 RL FEBS Lett. 345:93-98(1994).
 CC -1- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE
 IN HEMATOPOIETIC DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- TISSUE SPECIFICITY: LIVER AND PROMYELOCYTIC LEUKEMIA CELL
 LINE HL60.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING HEMATOPOIESIS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X67235; CAA47661.1; -;
 DR EMBL: L16499; AA02988.1; -;
 DR EMBL: Z21533; CAA79730.1; -;
 DR PIR: S26799; S26799.
 DR HSSP: P23441; 1FTT.
 DR TRANSFAC: T02092; -;
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFAM: PF00046; homeobox; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT DOMAIN 1 133 PRO-RICH.
 FT DNA_BIND 137 196 HOMEBOX.
 FT CONFLICT 115 115 L -> V (IN REF. 2).
 FT SEQUENCE 270 AA; 30021 MW; 453E9E6A CRC32;
 HMHP_HUMAN Length: 270 February 14, 2000 08:02 Type: P Check: 3785 ..
 1 MGFHPGPAA GAVGVPLYP TPLLOPAHPT PFYIDILGR GPAAPTPAP
 51 LSPNSNFTS LVSPYRTPVY EPTPIHAPV HHSAAALAA YPGGFGGPGP
 101 YPPRIVND YTHALRHDP LKPLWSPF QRLPKRRGG QVAFSNDQT
 151 ELEKKEETQ YLSPERKRL AKMLQSERQ YKTMQNRR AKMRLQENP
 201 QSNKKEELS LDSCDQROD LPSEONKGA LDSCSCSPS ASQEDLESEI
 251 SEDSDQVDI EGDKSYFNAG
 !!AA_SEQUENCE 1.0
 ID HMHP_MOUSE STANDARD: PRT; 271 AA.
 AC P43120;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HOMEBOX PROTEIN PRH (HOMEBOX PROTEIN HEX).
 GN PRHX OR HEX.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN: C57BL/6J; TISSUE: EMBRYO;
 RX MEDLINE: 93219088.
 RA BEDFORD F.K., ASHWORTH A., ENVER T., WIEDEMANN L.M.;
 RT "Hex: a novel homeobox gene expressed during haematopoiesis and
 conserved between mouse and human.";
 RL Nucleic Acids Res 21:1245-1249(1993).
 CC -1- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE
 IN HEMATOPOIETIC DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING HEMATOPOIESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z21524; CAA79729.1; -;
 DR HSSP: P23441; 1FTT.
 DR MGD: MGI:96086; HEX.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFAM: PF00046; homeobox; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT DOMAIN 1 134 PRO-RICH.
 FT DNA_BIND 138 197 HOMEBOX.
 FT SEQUENCE 271 AA; 29966 MW; 96744BE3 CRC32;
 HMHP_MOUSE Length: 271 February 14, 2000 08:02 Type: P Check: 8616 ..
 1 MGFHPGPAA GAVGVPLYP TPLLOPAHPT PFYIDILGR GPAAPTPAP
 51 TLSPNSNFT SLVSYRTPVY EPTPIHAPV SHHRAALAA AYGSFGGPGP
 101 YPPRIVND YTHALRHDP LKPLWSPF LQRLPKRRGG QVAFSNDQT
 151 VELEKKEETQ YLSPERKRL AKMLQSERQ YKTMQNRR AKMRLQENP
 201 POSNRKDALD SLDTSCQGO DLPSEONKGA SLDSCSCSPS PASQEDPQSE
 251 IESDSDQVD IEGDKYFNAG
 !!AA_SEQUENCE 1.0
 ID HPRK_BACSU STANDARD: PRT; 309 AA.
 AC O34483;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HPR(SER) KINASE/PHOSPHATASE (EC 2.7.1.-) (EC 3.1.3.-).
 GN PTK.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN
 RP SEQUENCE FROM N.A.
 RA LAZAREVIC V., SORDO B., RIVOLTA C., REYNOLDS S., MAUEL C.,
 RA KARAKATA D.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 RN [1]
 RP CHARACTERIZATION, AND SEQUENCE OF 1-10.
 RX MEDLINE: 98230327.
 RA REITZER J., HOISCHEN C., TITTEMEYER F., RIVOLTA C., RABUS R.,
 RA STELKE J., KARAWAYA D., SAIER M.H. JR., HILLEN W.;
 RT "A novel protein kinase that controls carbon catabolite repression in
 bacteria.";
 RL Mol. Microbiol. 27:1157-1169(1998).
 RN [3]
 RP CHARACTERIZATION OF THE PHOSPHATASE ACTIVITY.
 RX MEDLINE: 99141583.
 RA KRAVANJA M., ENGELMANN R., DOSSONNET V., BLUGGEL M., MEYER H.E.,
 RA FRANK R., GALINIER A., DEUTSCHER J., SCHNELL N., HENGSTENBERG W.;
 RT "The hprk gene of Enterococcus faecalis encodes a novel bifunctional
 enzyme: the hpr kinase/phosphatase.";
 RL Mol. Microbiol. 31:59-66(1999).
 CC -1- FUNCTION: THIS KINASE/PHOSPHATASE REGULATES CARBOHYDRATE UPTAKE
 AND METABOLISM BY PHOSPHORYLATING/DEPHOSPHORYLATING SER-45 OF THE
 PHOSPHOTRANSFER CARRIER PROTEIN (HPR) OF THE PHOSPHOTRANSFER-
 DEPENDENT SUGAR PHOSPHOTRANSFER SYSTEM (PTS) IN RESPONSE TO
 CYTOPLASMIC METABOLITE LEVELS AS A DEVICE FOR THE MODULATION OF
 CARBON CATABOLITE REPRESSION.
 CC -1- ENZYME REGULATION: ALLOSTERICALLY ACTIVATED BY METABOLITES SUCH AS

CC FRUCTOSE-1,6-BISPHOSPHATE AND INHIBITED BY INORGANIC PHOSPHATE.
 CC MASS SPECTROMETRY: MW=34529; METHOD=MALDI.
 CC -1- SIMILARITY: BELONGS TO THE PTKS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF017113; AAC67286.1; -
 CC DR EMBL: 299121; CAB15505.1; -
 CC DR SUBMITTER: BG14125; PTKS.
 CC KW Multifunctional enzyme; transferase; Serine/threonine-protein kinase;
 CC ATP-binding; Hydrolase.
 CC FT INIT MET 147 0
 CC FT NP_BIND 152 154 ATP (POTENTIAL).
 CC FT NP_BIND 159 159 ATP (POTENTIAL).
 CC SQ SEQUENCE 309 AA; 34571 MW; 3272EC0E CRC32;
 HPRK_BACSV Length: 309 February 14, 2000 08:02 Type: P Check: 4393 ..
 1 AAYRKDVE QPNLEISGE EGINRPTMS DLSRPGIEIA GYFTYPRER
 51 VOLLGTELS FEQLPEEEK KQMDSLCTD VTPAILSRD MPFQELIDA
 101 SERNGVPLR SPLKTRLLS RLTFPLESRL APTTAHGVV VDIYGVLLI
 151 TGRSGVQSE TALELVKRGH RLVAADCEVI ROEDQDLVNG NABELEHLL
 201 EIRGLIINV MLEFGAGAVR SNKRITIVAN LEWEOGROY DRGLSEETM
 251 KIIDEITKL TIIVPGRNL AVLEIYAAN FRLRMGLNA AEGFTKMLD
 301 VIEDEQEE
 11AA_SEQUENCE 1.0
 ID HUPK_RHLIV STANDARD; PRT; 370 AA.
 AC P28153;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYDROGENASE EXPRESSION/FORMATION PROTEIN HUPK.
 GN HUPK.
 OS Rhizobium leguminosarum (biovar victoriae).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-128C3;
 RA MEDLINE: 93108466.
 RA KEY L., HIDALGO E., PALACIOS J.M., RUIZ-ARGUESO T.;
 RT "Nucleotide sequence and organization of an H2-uptake gene cluster
 RT from Rhizobium leguminosarum bv. viciae containing a rubredoxin-like
 RT gene and four additional open reading frames.";
 RL J. Mol. Biol. 228:998-1002(1992).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 95020662.
 RA IMERIAL J., REV L., RUIZ-ARGUESO T.;
 RT "HUPK, a hydrogenase-ancillary protein from Rhizobium leguminosarum,
 RT shares structural motifs with the large subunit of NIFE hydrogenases
 RT and could be a scaffolding protein for hydrogenase metal cofactor
 RT assembly.";
 RL Mol. Microbiol. 9:1305-1306(1993).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B10;
 RA SCHMITT H., KOKOTER W., THUERING H., KERL V., BAUER T., FUCHS D.,
 RA TICH H., LOTZ W.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE HUPK FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X52974; CAA37158.1; -
 CC DR EMBL: 236981; CAA85440.1; -
 CC FT CONFLICT 240 240 A -> R (IN REF. 1).
 CC SQ SEQUENCE 370 AA; 38751 MW; 3F9E2409 CRC32;
 HUPK_RHLIV Length: 370 February 14, 2000 08:02 Type: P Check: 1548 ..
 1 MFLGAGTI GIDVTYSRAL ACSVAVKANR PGLTRMEVG ROPEEAPVLA
 51 GQVSLGFA QSVARLAVL AADIAWNE ERLGASGLL AERFETLRA
 101 LILQPTPLP ERFADAGRH LRELAASLA ISHKAART SRRLAAAE
 151 RLAAATLIG IPRGDTPLP ETACALIRH VEDHVFAGR RPDPLTISD
 201 AEVVARLDE AGYASLPHLS GRIAETGAYA RSASGLPEA PHLARLRAR
 251 IGVRLISLQ LTLARTGDF DCASLACGP TPGAGYICAV ECARGRLHQ
 301 IEIGSGRLA AVRILAPTEW NFRPAPPEVE TLSSPVGAD EAAVRSISRL
 351 AVLEPCVAF EIVREADA
 11AA_SEQUENCE 1.0
 ID HYCD_ECOLI STANDARD; PRT; 307 AA.
 AC B16430;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FORMATE HYDROGENLYASE SUBUNIT 4 (FHL SUBUNIT 4) (HYDROGENASE 3
 DE COMPONENT D).
 GN HYCD OR HEVD.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MC4100;
 RA MEDLINE: 90251163.
 RA BOEHM R., SAUTER M., BOECK A.;
 RT "Nucleotide sequence and expression of an operon in Escherichia coli
 RT coding for formate hydrogenlyase components.";
 RL Mol. Microbiol. 4:231-243(1990).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MC1655;
 RA BLATTNER F.R., BLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- PATHWAY: HYDROGEN METABOLISM; FHL PATHWAY.
 CC -1- SUBUNIT: FHL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE
 CC FROM FORMATE ARE RELEASED.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X17506; CA35549.1; -
 CC EMBL: U29579; AA69322.1; -
 CC EMBL: AE000356; AAC75764.1; -
 CC PIR: S08622; S08622.
 CC ECGENE: E610477; HYCD.
 CC PROSITE: PS00667; COMPLEX1_ND1_1; 1.
 CC PROSITE: PS00668; COMPLEX1_ND1_2; 1.
 CC PFM: PF00146; NADhdh; 1.
 CC Oxidoreductase; Transmembrane; Inner membrane.
 KW TRANSMEM 3
 FT TRANSMEM 68
 FT TRANSMEM 94
 FT TRANSMEM 132
 FT TRANSMEM 168
 FT TRANSMEM 222
 FT TRANSMEM 254
 FT TRANSMEM 285
 SQ SEQUENCE 307 AA; 33029 MW; 887A8868 CRC32;
 HYCD_ECOLI Length: 307 February 14, 2000 08:02 Type: P Check: 675 ..

1 MSVLYPLDIA LVLFVAVAPL SGITRVARAR LNRKPGVL QEYDIITKL
 51 GROSVPDAS GWFRITRYV MGVMLTAT ALPVTVGSP LPQIGDITL
 101 LVFAIARFE FAISGLDTGS PFTAIASRE AMGVAVPEM LLLGLWYAAQ
 151 VAGSTINISNI TDVYHWPIS QSIPLVIALC ACAFAITFEM GKLPFDIAEA
 201 EDELQEGPLS EYSGSGFGVM KMGISIKOLV VLQMFVGVEI PWGMETFTA
 251 GGLLALVIA IYKLVGVLY IALFENSMAR LRDIITPRIT WAGGFAPLA
 301 FVSLAA

!!AA_SEQUENCE 1.0
 ID IAA1_WHEAT STANDARD; PRT; 124 AA.
 AC P01085;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE ALPHA-AMYLASE INHIBITOR 0.19 (0.19 AI) (0.19 ALPHA-AI).
 OS Triticum aestivum (wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE: 85175148.
 RA MAEDA K., WAKABAYASHI S., MATSUBARA H.;
 RT "Complete amino acid sequence of an alpha-amylase inhibitor in wheat
 RT kernel (0.19-inhibitor).";
 RL Biochim. Biophys. Acta 828:213-221(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERNEL;
 RA OKUDA M., SATOH T., SAKURAI N., SHIBUYA K., KATI H., SAMEJIMA T.;
 RT "Overexpression in Escherichia coli of chemically synthesized gene for
 RT active 0.19 alpha-amylase inhibitor from wheat kernel.";
 RL J. Biochem. 123:918-926(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).
 RX MEDLINE: 98022743.
 RA ODA Y., MATSUNAGA T., FUKUYAMA K., MIYAZAKI T., MORIMOTO T.;
 RT "Tertiary and quaternary structures of 0.19 alpha-amylase inhibitor
 RT from wheat kernel determined by X-ray analysis at 2.06-A
 RT resolution.";

Biochemistry 36:13503-13511(1997).
 CC -1- FUNCTION: ALPHA-AMYLASE INHIBITOR.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- TISSUE SPECIFICITY: ENDOSPERM.
 CC -1- PM: THE DISULFIDE BONDS ARE ESSENTIAL FOR THE INHIBITOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AB003682; BAA20139.1; -
 CC PIR: A01324; WINTAL.
 CC PDB: 1HSS; 01-JUL-98.
 CC PROSITE: PS00426; CEREAL_TRYP_AMYL_INH; 1.
 CC PFM: PF00234; tryp_alpha_aml; 1.
 CC Alpha-amylase inhibitor; Multigene family; Seed: 3D-structure.
 KW DISULFID 6
 FT DISULFID 20
 FT DISULFID 41
 FT DISULFID 52
 FT DISULFID 6
 FT DISULFID 20
 FT DISULFID 28
 FT DISULFID 42
 FT DISULFID 54
 FT DISULFID 115
 SQ SEQUENCE 124 AA; 13337 MW; 05CA1FSD CRC32;
 IAA1_WHEAT Length: 124 February 14, 2000 08:02 Type: P Check: 3320 ..

1 SGPWMCYRQO AFQVPLPAC RELRLQNG SQVEANLRO CQQLAHISE
 51 WCRGALISM LDSMKRHEGA QEQAGTCAF PRCRREYVKL TAASITAVCR
 101 LPIVVASGD GAYYKDYAA YPDA

!!AA_SEQUENCE 1.0
 ID IAA1_WHEAT STANDARD; PRT; 123 AA.
 AC P01083;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE ALPHA-AMYLASE INHIBITOR 0.28 (CIII) (WMAI-1).
 OS Triticum aestivum (wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE.
 RA KASHIAN N., RICHARDSON M.;
 RT "The complete amino acid sequence of a major wheat protein inhibitor
 RT of alpha-amylase.";
 RL Phytochemistry 20:1781-1784(1981).
 RN [2]
 RP SEQUENCE, AND DISULFIDE BONDS.
 RX STRAIN-CV. PASTORE;
 RC MEDLINE: 91330910.
 RA POERIO E., CAPORALE C., CARRARO L., PUCCI P., BUONOCORE V.;
 RT "Assignment of the five disulfide bridges in an alpha amylase
 RT inhibitor from wheat kernel by fast-atom-bombardment mass
 RT spectrometry and Edman degradation.";
 RL Eur. J. Biochem. 199:595-600(1991).
 RN [3]
 RP MUTAGENESIS.
 RX MEDLINE: 92032760.
 RA GARCIA-MAROTO F., CARBONERO P., GARCIA-OLMEDO F.;
 RT "Site-directed mutagenesis and expression in Escherichia coli of
 RT WMAI-1, a wheat monomeric inhibitor of insect alpha-amylase.";
 RL Plant Mol. Biol. 17:1005-1011(1991).
 CC -1- FUNCTION: ALPHA-AMYLASE INHIBITOR.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: ENDOSPERM.

CC -1- PTM: THE DISULFIDE BONDS ARE ESSENTIAL FOR THE INHIBITOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR
 CC FAMILY.
 DR PIR: A01322; WMTA.
 DR HSSP: P01085; 1HSS.
 DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH. 1.
 DR PFAM: PF00234; TRYF_alpha_aml; 1.
 KW Alpha-amylase inhibitor; Multigene family; seed.
 FT DISULFID 7 54
 FT DISULFID 21 42
 FT DISULFID 29 82
 FT DISULFID 43 98
 FT DISULFID 56 113
 FT VARIANT 65 62
 FT VARIANT 67 67
 FT VARIANT 98 99
 FT CONFLICT 118 118 A -> A (IN VERY SMALL AMOUNT).
 FT CONFLICT 118 118 A -> C (IN REF. 1).
 SQ SEQUENCE 123 AA; 13326 MW; 82F56DD7 CRC32;
 IAA2.WHEAT Length: 123 February 14, 2000 08:02 Type: P Check: 4825 ..
 1 SGPWSCNCPA TGYKVSALTG CRAWKLOCV GSQVPEAVLR DCCQQLADIN
 51 NEMCRGDLG SMURSYQEL GVREGEVLP GCRREVKLT AASVEVCY
 101 PIPNSGDRA GVCYGDVAAY PDV
 IIAA-SEQUENCE 1.0 STANDARD; PRT; 124 AA.
 ID IAA5.WHEAT
 AC P01084;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE ALPHA-AMYLASE INHIBITOR 0.53.
 OS Triticum aestivum (wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Euphyllophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE: 83127436.
 RA MADDA K., HASE T., MATSUBARA H.;
 RT "Complete amino acid sequence of an alpha-amylase inhibitor in wheat
 RT kernel.";
 RL Biochim. Biophys. Acta 743:52-57(1983).
 RN [2]
 RP REVISIONS TO 119-124.
 RX MEDLINE: 85175148.
 RA MADDA K., WAKABAYASHI S., MATSUBARA H.;
 RT "Complete amino acid sequence of an alpha-amylase inhibitor in wheat
 RT kernel (0.19-inhibitor).";
 RL Biochim. Biophys. Acta 828:213-221(1985).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE: 84061717.
 RA MADDA K., WAKABAYASHI S., MATSUBARA H.;
 RT "Disulfide bridges in an alpha amylase inhibitor from wheat kernel.";
 RL J. Biochem. 94:863-870(1983).
 CC -1- FUNCTION: ALPHA-AMYLASE INHIBITOR.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- TISSUE SPECIFICITY: ENDOSPERM.
 CC -1- PTM: THE DISULFIDE BONDS ARE ESSENTIAL FOR THE INHIBITOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR
 CC FAMILY.
 DR PIR: A01323; WMTAS.
 DR HSSP: P01085; 1HSS.
 DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH. 1.
 DR PFAM: PF00234; TRYF_alpha_aml; 1.
 KW Alpha-amylase inhibitor; Multigene family; seed.
 FT DISULFID 6 115
 FT DISULFID 20 41 OR 20-42.
 FT DISULFID 28 83
 FT DISULFID 42 99 OR 41-99.
 FT DISULFID 42 99

SQ SEQUENCE 124 AA; 13185 MW; 3EE87339 CRC32;
 IAA5.WHEAT Length: 124 February 14, 2000 08:02 Type: P Check: 4073 ..
 1 SGPWSCNCPG AFQVPALPGC RPLIKLCNG SQVPEAVLRD CCQQLADISE
 51 WPRGALYSM LDMKKEHGV SEQAGTGAF PCRREVKLT TASTAVCR
 101 LPTVDSGD GAVCKDVAA YPDA
 IIAA-SEQUENCE 1.0 STANDARD; PRT; 312 AA.
 ID ICD2.RAT
 AC P52215; 035398;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CASPASE-2 PRECURSOR (EC 3.4.22.-) (CASP-2) (ICH-1 PROTEASE)
 DE (FRAGMENT).
 GN CASP2 OR ICH1.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP SEQUENCE OF 1-257 FROM N.A.
 RC TISSUE-KIDNEY CORTEX.
 RX MEDLINE: 98191309.
 RA KAUSHAL G.P., SINGH A.B., SHAH S.V.;
 RT "Identification of gene family of caspases in rat kidney and altered
 RT expression in ischemia-reperfusion injury.";
 RL Am. J. Physiol. 274:F587-F595(1998).
 RN [2]
 RP SEQUENCE OF 131-312 FROM N.A.
 RC TISSUE-OVARY.
 RX MEDLINE: 96042508.
 RA FLAHS J.A., KUGU K., TREBOVICH A.M., DESANTI A., TILLY K.I.,
 RA HIRSHFIELD A.N., TILLY J.L.;
 RT "Interleukin-1 beta-converting enzyme-related proteases (IRPs) and
 RT mammalian cell death: dissociation of IRP-induced oligonucleosomal
 RT endonuclease activity from morphological apoptosis in granulosa cells
 RT of the ovarian follicle.";
 RL Endocrinology 136:5042-5053(1995).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION. MIGHT FUNCTION BY EITHER
 CC ACTIVATING SOME PROTEINS REQUIRED FOR CELL DEATH OR INACTIVATING
 CC PROTEINS NECESSARY FOR CELL SURVIVAL (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF A SMALL AND A LARGE SUBUNIT (BY
 CC SIMILARITY).
 CC -1- PTM: THE MATURE PROTEASE CAN PROCESS ITS OWN PROPEPTIDE, BUT NOT
 CC THAT OF OTHER CASPASES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF025671; AAB82567.1;
 DR EMBL: U34684; AAC52260.1;
 DR HSSP: P29466; 1IBC.
 DR PROSITE: PS01121; CASPASE_HIS. 1.
 DR PROSITE: PS01122; CASPASE_CYS. 1.
 DR PFAM: PF00619; CARD. 1.
 DR PFAM: PF00655; ICE_P10. 1.
 DR PFAM: PF00656; ICE_P20. 1.
 KW Hydrolyase; Thiol protease; Apoptosis; Zymogen.
 FT NON_TER 1
 FT CHAIN 1 266 CASPASE-2 SUBUNIT 1 (BY SIMILARITY).
 FT CHAIN 267 312 CASPASE-2 SUBUNIT 2 (BY SIMILARITY).
 FT ACT_SITE 210 210 BY SIMILARITY.

FT ACT SITE 253 253 BY SIMILARITY.
RT NON PER 312 312
SQ SEQUENCE 312 AA: 35070 MW: 532B7699 CRC32:
ICE2.RAT Length: 312 February 14, 2000 08:02 Type: P Check: 8827

1 TLEKRELIQA KGGSESONE LNLNPKRGP QAFDACEAL RETQGHIED
51 LLTTLSDIQ HILPISCDY AQVSPSRQMS PALLISSPAY POILMNPYI
101 MMYVLPEFRW SHALLSTST LQLAIRLQSR PRGLALVSN VHTGEKDE
151 FRSGGDVDT TLVLEKILG YNVEHYLDQT AQEMOEKLN FAQIPAHRT
201 DSCIVALLSH GVEGGIYVD GKLLQLOEVF RLEDNANCPN LONKRMFEI
251 QACRGDETR GVDODGKNH AQPGCESD TYKEELMKR LPTSDMICV
301 YACIKDNAPI RN

11AA_SEQUENCE 1.0 STANDARD: PRT: 207 AA.
ID ICW3_PSCORE
AC P10822:
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE CHYMOTRYPSIN INHIBITOR 3 PRECURSOR (WCI-3)
OS PSEPHOCARPUS TETRAGONOLOBUS (Gosa bean) (Asparagaceae)
OC Euparieta, Vitidiplanetae, Streptophyta, Embryophyta, Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta, eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Psophocarpus.
RN [1]
RP SEQUENCE FROM N.A.
RA PEYACHOKNAGUL S., MATSUI T., SHIBATA H., HARA S., IKENAKA T.,
RA OKADA Y., OHNO T.;
RT "Sequence and expression of the mRNA encoding the chymotrypsin
RT inhibitor in winged bean (Psophocarpus tetragonolobus (L.) DC.).";
RL Plant Mol. Biol. 12:51-58(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA HAHU Y., PEYACHOKNAGUL S., UMEMOTO K., SAKATA Y., OHNO T.;
RA MEDLINE: 9224991.
RT "Structure and regulated expression of Kunitz chymotrypsin inhibitor
RT genes in winged bean [Psophocarpus tetragonolobus (L.) DC.]";
RL J. Biochem. 111:249-258(1992).
RN [3]
RP SEQUENCE FROM N.A.
RA HAHU Y., SAKATA Y., FUKASAWA K., OHNO T.;
RA MEDLINE: 94122379.
RT "Ubiquitous nuclear proteins bind to 5' upstream region of major
RT Kunitz chymotrypsin inhibitor gene in winged bean";
RL Plant Mol. Biol. 23:1139-1150(1993).
RN [4]
RP SEQUENCE OF 25-207.
RA SHIBATA H., HARA S., IKENAKA T.;
RT "Amino acid sequence of winged bean (Psophocarpus tetragonolobus (L.)
RT DC.) chymotrypsin inhibitor, WCI-3.";
RL J. Biochem. 104:537-543(1988).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.95 ANGSTROMS).
RC TISSUE-SEED:
RA DATAGUPTA J. K., PODDER A., CHAKRABARTI C., SEN U., DUTTA S. K.,
RA SINGH M.;
RT "Structure of a Kunitz-Type chymotrypsin from winged bean seeds at
RT 2.95-A resolution.";
RL Acta Crystallogr. D 52:521-528(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RA MEDLINE: 99258819
RA DATAGUPTA J. K., PODDER A., CHAKRABARTI C., SEN U., MUKHOPADHYAY D.,
RA DUTTA S. K., SINGH M.;

RT RT "Refined crystal structure (2.3 Å) of a double-headed winged bean
RT alpha-chymotrypsin inhibitor and location of its second reactive
RT site.";
RL Proteins 35:321-331(1999).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.13 ANGSTROMS).
RA RAVICHANDRAN S., SEN U., CHAKRABARTI C., DATAGUPTA J. K., DUTTA S. K.,
RA SINGH M.;
RT "2.13-Å structure of a Kunitz-type winged bean chymotrypsin inhibitor
RT protein." (OCT-1998) to the PDB data bank.
RL Submitted:
CC -1- FUNCTION: INHIBITS ALPHA-CHYMOTRYPSIN AT THE MOLAR RATIO OF 1:2
CC IN STATE OF 1:1.
CC -1- SIMILARITY: TO OTHER LEGUMINOUS KUNITZ-TYPE INHIBITORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S6732; AAC60535.1;
DR EMBL: D13974; BAA03084.1;
DR EMBL: D13975; BAA03085.1;
DR EMBL: D13976; BAA03086.1;
DR PIR: JX0206; JX0206.
DR PIR: JX0206; JX0206.
DR PIR: S42563; S42563.
DR PIR: S42564; S42564.
DR PDB: 1WBC; 03-APR-96
DR PDB: 2WBC; 25-FEB-98
DR PDB: 4WBC; 12-MAR-99.
DR PROSITE: PS00283; SOYBEAN_KUNITZ_1.
DR PFM: PF00197; Kunitz_Legume; 1.
DR Serine protease inhibitor; Signal: 3D-structure.
FT SGNAL 1 24
FT CHAIN 25 207 CHYMOTRYPSIN INHIBITOR 3.
FT DISULFID 65 109
FT DISULFID 159 168
FT ACT_SITE 89 90 REACTIVE BOND (CHYMOTRYPSIN).
SQ SEQUENCE 207 AA: 22774 MW: 2288B770 CRC32:
ICW3_PSCORE Length: 207 February 14, 2000 08:02 Type: P Check: 7697

1 MKSTFLALF LLSALISHLP SSTADDLVD AEGNLVENG TYLLPHIMA
51 HGGIEIETAKI GNEPCPLTV RSPNEVSKGE PIRISSOFLS LPIRGSIVA
101 LGFANPPSCA ASPWTAVDS PGQPAVKLSQ QKLPEKDIIV EFKFVSHSN
151 IHVYKLLYCO HHEEDVKCDO YIGIHDRNG NRRLVVTEEN PLEIVILKAK
201 SETASSH

11AA_SEQUENCE 1.0 STANDARD: PRT: 203 AA.
ID IDLACTV
AC P7202; 008150.
DT 13-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE POTATIVE ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (EC 5.3.3.2) (IPP
DE ISOMERASE) (ISOPENTENYL PYROPHOSPHATE ISOMERASE).
GN RV1745C OR MTCY28.08C OR MTCY04C12.29C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-H37RV;
RX MEDLINE: 98295987.
RA COLE S. T., BROSC R. J., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S. V., EIGLMEIER K., GAS S., BARRY C. E. III, TEKAIA F.,

RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNBY T., JAGELS K., KROCH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE J., OUALI M.A., RAJANDREAN M.A., ROGERS J.,
 RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -1- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
 CC HOMALYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS ALLYLIC ISOMER,
 CC DIMETHYLLYL DIPHOSPHATE (DMAPP) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ISOPENTENYL DIPHOSPHATE - DIMETHYLLYL
 CC DIPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE IPP ISOMERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: 295890; CAB09331.1; -
 KW Hypothetical protein; Isomerase; Isoprene biosynthesis.
 FT ACT_SITE 75 BY SIMILARITY.
 FT ACT_SITE 124 BY SIMILARITY.
 SQ SEQUENCE 203 AA; 22489 MW; 4DAFD37 CRC32;
 IDL_MYCTU Length: 203 February 14, 2000 08:02 Type: P Check: 9024 ..
 1 MRSYRPAAP IERYVLNDR GDAFGVADKA TVHTGDTPLH LARSSYFDL
 51 HQGLITRRA ATKRTPAVM TNSCGHPLE GSELPGAIR RLAAELGLTP
 101 DRYDLIPGF RYRAMADGT VENEICPYR VQVDQPRPN SDEVDAIRWL
 151 SMEQFVRDVT AGVAPVSPW CRSQGLYLR LQPCRAQMPV ADCCRLPKAA
 201 HGN
 !!AA SEQUENCE 1.0
 ID IDNR_ECOLI STANDARD: PRT; 332 AA.
 AC P39343;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE L-IDONATE REGULATORY PROTEIN.
 GN IDNR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 953434362.
 RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
 RA BLATTNER F.R.,
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes."
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [2]
 RP FUNCTION: 98324983.
 RA BAUSCH C., PEERKHANUS N., UTZ C., BLAIS T., MURRAY E., LOWARY T.,
 RA CONWAY T.,
 RT "Sequence analysis of the GntII (subsidiary) system for gluconate
 RT metabolism reveals a novel pathway for L-idonic acid catabolism in
 RT Escherichia coli."
 RL J. Bacteriol. 180:3704-3710(1998).
 CC -1- FUNCTION: IDN OPERON REGULATOR. MAY REPRESS GNTKU AND GNTT GENES
 CC WHEN GROWING ON L-IDONATE.

CC -1- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U14003; AA97161.1; -
 DR EMBL: AE000497; AAC77221.1; -
 DR HSSP: P03023.1; ICD.
 DR ECOGENE: EG12538; IDNR.
 DR PROSITE: PS00356; Iaci; 1.
 DR PFAM: PF00356; Iaci; 1.
 KW Transcription regulation; DNA-binding; Gluconate utilization.
 FT DNA_BIND 8 H-T-H MOTIF (POTENTIAL).
 FT SEQUENCE 332 AA; 37567 MW; B4EIF229 CRC32;
 IDNR_ECOLI Length: 332 February 14, 2000 08:02 Type: P Check: 2843 ..
 1 MRNRISLQD IATLAVTKV TVSRVIRSP KYAKETGERI AKIMEINYI
 51 PNRAPGMLN AGSYTLGILI PSFONLFAD ILAGIESVTS EHYQTILAN
 101 YNYRDSSEE SYVNLSTYNI DGIILSEKH TIRTVFLNS ATIVVELMD
 151 VQGRLEMEV GFQNRQAQFD MVTMLKRY RHIIILGSK DQTRDEQRYQ
 201 GYCAMMLHN LSPLRNRPRA ISSIHLGML MDALSNAND LDGFCINDD
 251 IANALLDLC ERYLAVPEOI SIAGFHGELI GRQMIFSLAS VITRFPIGR
 301 MAQMLLSKI KNDNHNTV DLYQIYHGN TL
 !!AA SEQUENCE 1.0
 ID IDNR_ECOLI STANDARD: PRT; 131 AA.
 AC P09182;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE COLICIN N IMMUNITY PROTEIN (MICROCIN N IMMUNITY PROTEIN).
 GN CNI.
 OS Escherichia coli.
 OC Plasmid Coln pCHAP4.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 88174431.
 RA PUGSLEY A.P.,
 RT "The immunity and lysis genes of Coln plasmid pCHAP4."
 RL Mol. Gen. Genet. 211:335-341(1988).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X06933; CAA30020.1; -
 DR PIR: S01761; IMECN4.
 KW Plasmid; Bacteriocin; Colicin; Transmembrane.
 FT TRANSMEM 66
 FT TRANSWEM 104 84
 FT SEQUENCE 131 AA; 15245 MW; BA839D94 CRC32;

RT system of Streptomyces albus G.;

RL Mcl. Gen. Genet. 225:142-147(1991).

CC -1- FUNCTION: INACTIVATION OF SAL I RESTRICTION-MODIFICATION SYSTEM.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X56644; CAA39965.1; -

KM Transposable element.

SO SEQUENCE 256 AA; 28409 MW; ADSCD17C CRC32;

IS12_STRAL Length: 256 February 14, 2000 08:02 Type: P Check: 4079 ..

1 MAGVITASEP SWIAPFSGLS PROPGKLVTV LRREGADAVR KGRFWSIPLE

51 DRALLVAAYV RNTLTMROLA PLFGVSKSAA DRIDHIGPM LALQPRKRA

101 KDTVLIVDGT LVPTRDHTIA ERSKNRYST NQVVIDADT RLVVVYGRPL

151 AGNRNCKAV EESGAKAAGV KTLIADGXY PGTGLVTPHR RENGQAGLPD

201 WKEHNKSHK QVAREVEHF ARMTWKILR DCLRLGDGVH HAMLGIARMH

251 NLAALTG

II.AA SEQUENCE 1.0

ID ISIA.SYNY3 STANDARD: PRT: 342 AA.

AC 055274; P73884; -

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE IRON-STRESS INDUCED CHLOROPHYLL-BINDING PROTEIN.

GN ISIA OR SL0247.

OS Synchocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.

RN [1]

RP SEQUENCE FROM N.A.

RA FERRIERA F., STRAUS N.A.;

RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,

RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,

RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NABURO K.,

RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,

RA YAMADA M., YASUDA M., TABATA S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synchocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res 3:109-136(1996).

CC -1- FUNCTION: PHOTOSYSTEM II CHLOROPHYLL-BINDING PROTEIN.

CC -1- INDUCTION: BY IRON STRESS.

CC -1- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L26530; AAA27291.1; -

DR EMBL: D90910; BAA17948.1; -

DR PFAM: PF00421; PSII; 2.

KW Photosystem II; Chlorophyll; Transmembrane.

FT TRANSMEM 30 51 1.

FT DOMAIN 63 81 II. VERY HYDROPHOBIC.

FT TRANSMEM 89 113 III.

FT TRANSMEM 141 167 IV.

FT TRANSMEM 204 223 V.

FT TRANSMEM 240 262 VI.

FT DOMAIN 302 323 VII. VERY HYDROPHOBIC.

FT CONFLICT 94 94 S -> R (IN REF. 1).

FT CONFLICT 334 342 EDAFDSLOT -> NKRLIPCKPSH (IN REF. 1).

SO SEQUENCE 342 AA; 37221 MW; 889359C7 CRC32;

ISIA.SYNY3 Length: 342 February 14, 2000 08:02 Type: P Check: 4153 ..

1 MOTYGNDTVQ YEWMAGNARF ADOSGLFIA HVAQAALTA F WAGATTFEI

51 SNFDTQAMG DOGLILPHL ATLGVGVGDG GQIVDTYPYF VIGSIHLIAS

101 AVLGAGALFH TLRAPADLST LKGQKKRHF TWENPOOLGI ILGHLLPLG

151 AGALLAGKA MYWGLYDAT TQTVLVSOP TLDPVITYG QTHFASISL

201 EDLVGHTIFV GFLIGGIM HILVPLGMA KVLLESSEA ILSYSLGIA

251 LAGFAAYFC AVNTLAPPE FYGPPIAKL GIPFPADV ELPNHAHSR

301 AMLNANHEFL APFLQGHM HALRALGDF KRYEQAFDSL QT

II.AA SEQUENCE 1.0

ID KDD.D.BACSU STANDARD: PRT: 254 AA.

AC P50842; -

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 2-DEOXY-D-GLUCONATE 3-DEHYDROGENASE (EC 1.1.1.125) (2-KETO-3-DEOXYGLUCONATE OXIDOREDUCTASE).

GN KDD.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-168 / MARBURG;

RX MEDLINE: 96349105.

RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,

RA SERROR P.;

RT "Sequence analysis of the Bacillus subtilis chromosome region between

RT the sera and kgd loci cloned in a yeast artificial chromosome.";

RL Microbiology 142:2005-2016(1996).

CC -1- CATALYTIC ACTIVITY: 2-DEOXY-D-GLUCONATE + NAD(+) -> 3-DEHYDRO-2-DEOXY-D-GLUCONATE + NADH.

CC -1- PATHWAY: PEPTIN DEGRADATION.

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L47838; AAB38476.1; -

DR EMBL: Z99115; CAB14131.1; -

DR HSBP; P19992; 2HSD.

DR SUBTILIST; BG11400; KDD.

DR PROSITE; PS00061; ADH.SHORT.1.

DR PFAM; PF00106; adh.short.1.

KW Oxidoreductase; NAD.

FT NP_BIND 16 40 NAD (BY SIMILARITY).

FT ACT SITE 159 159 BY SIMILARITY.

SO SEQUENCE 254 AA; 27185 MW; 44F5CB25 CRC32;

KDDJ_BACSU Length: 254 February 14, 2000 08:02 Type: P Check: 3484

1 MG1HDAFSL KGKTAIVTGP GTGIGIGIAK ALAGAGADII GTSHTSLSE
51 TOOLVQEGFR IFTSFILDMS KPEAIKDSAA ELFENROIDI LVNNAGIIHR
101 EKAEDPEEN WQVHLNVNLSL IFTIQLQAG RHMLKRGHGK IINIASLISF
151 OGGLIVPATV ASKHAVAGLT KSPANWMAAS GIOVVAIAG YISANTKPI
201 RODEKKNEDI LKRIPAGRW QADIDIGTAV FLASASDXY NGLIANDGS
251 WLSR

11AA_SEQUENCE 1.0 STANDARD: PRT: 467 AA.
ID KRL_HSVBS
AC 008097;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE (EC 2.7.1.-).
OS Bovine herpesvirus type 1.2 (strain ST).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94167875.
RA LEUNG-TRICK P., AUDONNET J.F., RIVIERE M.;
RT "The complete DNA sequence and the genetic organization of the short
RT unique region (US) of the bovine herpesvirus type 1 (ST strain)."
RL Virology 199:409-421(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 223068; CAAB0602.1;
DR HSSP: P24941; IAOI.
DR PROSITE: P500107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR PROSITE: P500108; PROTEIN_KINASE_ST; 1.
DR PROSITE: P50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; PKinase; 1.
KW transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 164 462 PROTEIN_KINASE.
FT NP_BIND 170 178 ATP (BY SIMILARITY).
FT BINDING 194 194 ATP (BY SIMILARITY).
FT ACT_SITE 279 279 BY SIMILARITY.
SO SEQUENCE 467 AA; 49885 MW; 25EB1E22 CRC32;

KRL_HSVBS Length: 467 February 14, 2000 08:02 Type: P Check: 5295

1 MERAERLRL QRAGLTMSR FACCVAFPS GSRLGSGVNG AAAPARCAA
51 EGSADILAV NNEGEVAPR ARTGPPDAG IEGGAAGVNG EQGVAAAGNE
101 RRAATGDERE SASGENSESE SESSESESE SGADGDMD DDDAGPAGV
151 TREBAGGAR ALNFRIIRL TPGEGRVFE ATGPAPAGH VVAKIGASAS
201 TLAEMILRT LDHANVVLK AVLFHGLVC AVIARREDL HTLMKIDRP
251 MALPTALQVT RAVLOGLAIV HSRRIARHV KTEVFLNGP GDVCLIDFGA
301 AHGPVTEPR YGLAGTLETN SPILLARARY DCRDVMASG VVAYEMLAYP
351 RALFDSBAPG OGEDAASGP PTLIGRDCA ROLLRYIRL AVHAEEPPPS
401 PTDRLTRNF RHAATGREPH SPYCLAVLR LPCDADRLH QMLTEFERAR

451 PTAAELLEHP VEGAASG

11AA_SEQUENCE 1.0 STANDARD: PRT: 202 AA.
ID LSPA_MYCTU
AC 010764;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROBABLE LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36) (PROLIPOPROTEIN
DE SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II).
OS LSPA OR RV1539 OR MYCY48.26C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98265987.
RX STRAIN-H37RV;
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGMEIER K., GINS S., BARRY C.E. II, TERKLA E.,
RA BADCOCK K., BASHAM D., BROWN D., CHILINCWORTH H., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAUNDREEM M.A., ROGERS J.,
RA RUTHER S., SEEGER R., SKELTON S., SQUARES S., SQUARES R., SULLSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY CATALYZES THE REMOVAL OF
CC SIGNAL PEPTIDES FROM PROLIPOPROTEINS.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF N-TERMINAL LEADER SEQUENCES FROM
CC MEMBRANE PROLIPOPROTEINS. HYDROLYSES XAA-XBB-XBB-1-CYS, IN WHICH
CC XAA IS HYDROPHOBIC (PREFERABLY LEU), XBB IS OFTEN SER OR ALA, XCC
CC IS OFTEN GLY OR ALA, AND THE CYS IS ALKYLATED ON SULFUR WITH A
CC DICYCLIC CYCERYL GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 274020; CA98323.1;
DR PROSITE: P500855; SPASE_II; 1.
DR PFAM: PF01252; Peptidase_A8; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease; Transmembrane.
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 35 55 POTENTIAL.
FT DOMAIN 56 87 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 88 108 POTENTIAL.
FT DOMAIN 109 111 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 112 132 POTENTIAL.
FT DOMAIN 133 159 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 160 180 POTENTIAL.
FT DOMAIN 181 202 CYTOPLASMIC (POTENTIAL).
SO SEQUENCE 202 AA; 21345 MW; 906DB13F CRC32;

LSPA_MYCTU Length: 202 February 14, 2000 08:02 Type: P Check: 2319

1 MPDEPGSAD PLTTEBAGG AGEPRNAPR RLRLMLSLVA VVVLTIDIVT
51 KYVAVOLLP GOPVSIIGDT VVTTLVRNSG AAFSNATGYT WVLTLATGV
101 VVGIFWGR LVSPWMAIGL GMLLGAMGN LVDRFFRAG PLRGHVVDFL
151 SVGWPEFNV ADPSVGGAI LLYIISIGF DEDTVGRRA DSDTVGRRA
201 DG

11AA_SEQUENCE 1.0 STANDARD; PRT; 277 AA.
 ID MAUF_PARDE
 AC P29897;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE METHYLAMINE UTILIZATION PROTEIN MAUF.
 GN MAUF.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Paracoccus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 413;
 RX MEDLINE; 95045590.
 RA VAN SPANING R.J.M., DER PALEN C.J., SLOTBOOM D., REIJNDERS W.N.,
 RA SOUTHWAMER A.H., DUINE J.A.;
 RT "Expression of the mau genes involved in methylamine metabolism in
 RT Paracoccus denitrificans is under control of a LysR-type
 RT transcriptional activator."
 RL Eur. J. Biochem. 226:201-210(1994).
 RN [2]
 RP SEQUENCE OF 129-277 FROM N.A.
 RX MEDLINE; 92272712.
 RA CHRISTENSEN D.A.Y., BOYD J., MATHENS F.S., LIDSTROM M.E.;
 RT "A transcription-dependent DnaE I hypersensitive site in a far
 RT upstream segment of the chicken alpha-globin gene domain coincides
 RT with a matrix attachment region."
 RL Biochem. Biophys. Res. Commun. 184:1181-1189(1992).
 CC -1- PATHWAY: METHYLAMINE UTILIZATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U12464; AAA56723.1;
 CC EMBL: M90099; AAA25579.1;
 CC PIR: PH0856; PH0856.
 DR DR
 KW Transmembrane.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 237 277 POTENTIAL.
 SQ SEQUENCE 277 AA; 28361 MW; 7BE517A2 CRC32;
 MAUF_PARDE Length: 277 February 14, 2000 08:02 Type: P Check: 7474 ..

15-JUL-1998 (Rel. 36, Last annotation update)
 DE MALATE DEHYDROGENASE (EC 1.1.1.37).
 GN MDH.
 OS Rhizobium leguminosarum (biovar viciae).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-3841;
 RA POOLE P.S., ALLAWAY D., SMITH M.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-MALATE + NAD(+) = OXALOACETATE + NADH.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ002750; CA05717.1;
 CC DR HSP: Q27743; 1LDG.
 DR PROSITE: P500068; MDH; FALSE_NEG.
 DR PIR: PF00056; 1db; 1.
 KM Oxidoreductase; Tricarboxylic acid cycle; NAD.
 FT ACT_SITE 149 149 PROTON-RELAY (BY SIMILARITY).
 FT BINDING 152 152 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
 FT ACT_SITE 176 176 PROTON-RELAY (BY SIMILARITY).
 SQ SEQUENCE 320 AA; 33590 MW; 192E3E24 CRC32;
 MDH_RHIV Length: 320 February 14, 2000 08:02 Type: P Check: 1419 ..

1 MAANKIALIG SCMIGITLH LAGIKELGDI VLFDIADGIP QGKGLDISQS
 51 SPEGEDVNL TGASDYSIE GADVCIYTAG VARKGMSRD DLGINLKV
 101 EGVAGIKKY APNAVICIT NPIDAWMAL QKSGLPANK VGVNAGVLS
 151 SRRFLAKE FNVSVODVTA FVLGSHDYM VPLARYSTVG GIPLTDLVTM
 201 GWTKERLEE IIOFRDGA EIVGLKTGS AYAFAASAI EMASYLKDK
 251 KVLPCAAH SGQGVKDMY VGVPTVIGAG GVERIIEIDL NKTEKAFDK
 301 SVGVAVGLCE ACINIPALK
 11AA_SEQUENCE 1.0 STANDARD; PRT; 430 AA.
 ID MDH_RHIV
 AC P4689;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PUTATIVE CELL CYCLE PROTEIN MESJ HOMOLOG.
 GN MESJ OR H10404.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., CORTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN P.T., SAUDEK D.M., BRANNON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRGAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.,
 RT "Whole-genome random sequencing and assembly of Haemophilus

RT Influenzae Rd.":
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32723; AAC22063.1; -
 DR TIGR: W10404;
 DR SEQUENCE 430 AA; 50052 MW; BE205D2 CRC32;
 SQ
 MESJ_MAEIN Length: 430 February 14, 2000 08:02 Type: P Check: 1410 ..
 1 MDLSDIERO LOKATAQAFI IALSGGLDST VLISLFAKIC OKOHPHPLS
 51 IRAIHINHL SPNADSWAKH CODICDOFOI PLIERVOVD KTNCEIAGAR
 101 EARYQAIKKY IOTOEMLYTA HMLNDQETIF FLAKRGSGI KGLGMOQOS
 151 VLFQMLIRP LIGFTRTOLE NYAQEKILNW ITDESNEENR YDRNFLRNEI
 201 LPELRERMAH FDLAVORSNO HCFEODOLIN DLLSEIFTEH COIKNOERLC
 251 QFQYQSIARQ TALLRMWLAE NOLEMPKRRQ LTOLINDVIF AKERANPOFO
 301 LVNKVIRRYQ DSYLTKEPFS DLTKCTKLE QNTLNPDOL GNLFVQENEH
 351 NLFFWQDYS VLEKTNLPI SIRGYSGVK KHPKPRPD IKIMQELGV
 401 PPMERNRIP LFFGNEKSA VGFEFVRLKSS
 11AA SEQUENCE 1.0
 ID MPXA MYXA STANDARD; PRT; 236 AA.
 AC P38371;
 DT 01-OCT-1994 (rel. 30, Created)
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 01-FEB-1995 (rel. 31, Last annotation update)
 DE PUTATIVE LIPOPROTEIN MLPA PRECURSOR.
 GN MLPA.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 CC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DEFL.
 RA MARTINEZ-CANAMERO M., MUNOZ-DORADO J., FAREZ-VIDAL E., INOUE M.,
 RA INOUE S.;
 RT "Car. a 115-kilodalton membrane protein required for development of
 RT Myxococcus xanthus".
 RT J. Bacteriol. 175:4756-4763(1993).
 RL -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S64103; AA27615.1; -
 DR PIR: B40609; B40609.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR Membrane; Signal; Lipoprotein.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 236 PUTATIVE LIPOPROTEIN MLPA.

FT LIPID 22 22 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 236 AA; 24618 MW; 97CAB418 CRC32;
 MPXA MYXA Length: 236 February 14, 2000 08:02 Type: P Check: 8809 ..
 1 MKRNIVNTAL VLVAGSILT GCNFEQEPETN CFVQSPSWA VKYVDVDSRK
 51 DANGDECTTT APLVELMGV KYVNPETGAA QALARPATIA SRAIDTTT
 101 SADQTSIGSL DTEPRDGHFC HANDFAPEV NVAASDTAA NTRYEFETV
 151 RYSAAVAPG TQFGEKLT SNQCTSSYVA RAWPPAPCD TASTEPAENC
 201 GVGSGLNPEF AVVQPSAT GTGCVCPAG DIPSRK
 11AA SEQUENCE 1.0
 ID MML6 MYCTU STANDARD; PRT; 397 AA.
 AC Q10773;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-DEC-1999 (rel. 39, Last annotation update)
 DE PUTATIVE MEMBRANE PROTEIN MML6.
 GN MML6 OR RV1357 OR MCT148.08C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 98295987.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. II, TEKAIA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELLWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAN M.A., ROGERS J.,
 RA RUTHER S., SEEGER K., SKELTON S., SQUARES S., SULLSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence".
 RT Nature 393:537-544(1998).
 RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE MML6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z74020; CA98334.1; -
 DR KX Hypothetical protein; Transmembrane.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 210 210 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 SQ SEQUENCE 397 AA; 42421 MW; 9AFDFAD0 CRC32;
 MML6 MYCTU Length: 397 February 14, 2000 08:02 Type: P Check: 4418 ..
 1 MGIISTGVY KKGWVRVSF DTIDIGIQLG EQLASTVTL DKLAIQPOL
 51 VALLPEIAS QOINRELAIA NYATMSGIYA QTAALIEENA AMGAFDAK
 101 NDSFVLPPE AFDPDFORG LKFLSADGK AARMTISHEG DPATPEGIS
 151 IDAIKQANE AVKGTPMAGA GIYIAGTAAT FPDIDGATY DLLIAGTAAL
 201 SLILIMMI TRSLVAALVI VGIVALSLGA SFGSLVWQ HLLGIQIYWI

251 VLAVALLL AVSDYNLL ISREKEIGA GLNTGIIRAM AGTGVATTA
301 GVPFAAMSS FVPSDLRVG QIGTTIGGL LPTLYVRAF MTPSIVLGL
351 RMFWPQVR PRPASMRLR YGPRVREL LHEGDDER TOVALHR

11AA_SEQUENCE 1.0 STANDARD; PRT; 288 AA.

AC P55818;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE METHYLENETRAHYDROFOLATE DEHYDROGENASE [NAD+] (EC 1.5.1.15).
GN MTDA.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Methylobacterium.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-AMI / NCIB 9133;
RX MEDLINE: 94193542.
RA CHISTOSERDOVA L.V., LIDSTROM M.E.;
RT Identification of sgaa and mtda and sequences of sgaa, hpra, and
RT mtda.
RT J. Bacteriol. 176:1957-1968(1994).

CC -1- FUNCTION: MAY PLAY AN INTEGRAL ROLE IN METHYLOTROPHIC METABOLISM,
CC EITHER IN FORMALDEHYDE OXIDATION OR AS PART OF THE SERINE CYCLE.
CC -1- CATALYTIC ACTIVITY: 5,10-METHYLENETRAHYDROFOLATE + NAD(+) =
CC 5,10-METHYLENETRAHYDROFOLATE + NADH.
CC -1- PATHWAY: SERINE PATHWAY.
CC -1- SIMILARITY: ONLY DISTANTLY RELATED TO THE OTHER 5,10-METHYLENE-TH
CC DEHYDROGENASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC DR EMBL: L27235; NOT ANNOTATED_CDS.
CC KW Oxidoreductase; NAD: One-carbon metabolism.
CC SQ SEQUENCE 288 AA; 29736 MW; DECIBD20 CRC32;

MTDA_MREX Length: 288 February 14, 2000 08:02 Type: P Check: 9524

1 MSKLLLOFD TDAIPSVFV VGVYDGDH ITGVGNTPD NVGATVDDTI
51 YTRGKREKOS TAIIVGGDM AAGEVFEAV KKEFFGFRV SCMDSDNSN
101 TTAAGVALV VKAAGSVKG KKAIVLACTG PVGRSAALL AGECAEYVLC
151 GRKIDKQAA ADSVNRKRY NTALEIVDD ASRAEAVKA HEVFTAGLIG
201 LELLQAAMQ NESSIEIVD YNAOPPLIG GIDATDKKE YGKRAFQAL
251 GIGIKIKLH RACIAKLFS SEGVDAEEL YKLAKWA

11AA_SEQUENCE 1.0 STANDARD; PRT; 120 AA.

ID NEUB_XENLA
AC P43443;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROMEDIN B PRECURSOR.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae;
CC Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae;
CC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.

CC TISSUE-BRAIN;
CC MEDLINE: 95028554.
CC RA WECHSELBERGER C., KREIL G., RICHTER K.;
CC RT Isolation and sequence of a cDNA encoding the precursor of a
CC bombesin-like peptide from brain and early embryos of Xenopus
CC laevis.
CC RT Proc Natl Acad Sci U S A. 89:9819-9822(1992).
CC -1- FUNCTION: MAY BE A GROWTH FACTOR REQUIRED FOR THE HIGH RATE OF
CC CELL DIVISION ENSUING AFTER FERTILIZATION, OR MAY HAVE A ROLE AS A
CC SECRETAGOGUE THAT TRIGGERS THE RELEASE OF STORED COMPONENTS.
CC -1- TISSUE SPECIFICITY: BRAIN, INTESTINE, AND OVARIAS AND EARLY
CC EMBRYOS (STAGES 2 AND 10).
CC -1- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC DR EMBL: L01530; AAA49912.1;
CC DR PROSITE: PS00257; BOMBESIN: 1.
CC KW Bombesin family; Amidation; Cleavage on pair of basic residues;
CC SIGNAL.

FT SIGNAL. 1 29 POTENTIAL.
FT PROPEP 30 44
FT PEPTIDE 45 54 NEUROMEDIN B.
FT MOD RES 54 54 AMIDATION (G-55 PROVIDE AMIDE GROUP).
SQ SEQUENCE 120 AA; 14384 MW; F3870792 CRC32;

NEUB_XENLA Length: 120 February 14, 2000 08:02 Type: P Check: 1414

1 MSAVPLTRML PIRFTLLH LSTPLPYFCM ESEDAENIE KIRGNOMAI
51 GHEMKKSIQ DYNESODM DSEDFPRRI EIKRTFROE PIRALSPKO
101 DEIQMWIKRI MDVYIKTKQK

11AA_SEQUENCE 1.0 STANDARD; PRT; 229 AA.

ID NIFY_KLEPN
AC P09135;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE NIFY PROTEIN.
GN NIFY.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Klebsiella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89041575.
RA BERON J., CANNON M., BANNAN-WOLLASTON V., ALLY A., SUTTEROUIST R.,
RA CANNON F.;
RT The nucleotide sequence of the nify, nify, nifx and nify genes of K.
RT pneumoniae.
RL Nucleic Acids Res. 16:9860-9860(1988).

CC SEQUENCE FROM N.A.
RX MEDLINE: 89094839.
RA ARNOUD W., RUMP A., KLIPP W., PRIEFER U.B., PUHLER A.;
RT Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the
RT entire nitrogen fixation gene cluster of Klebsiella pneumoniae.
RL J. Mol. Biol. 203:715-738(1988).
CC -1- SIMILARITY: TO NIFY PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X75772; CA53393.1; -
 CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 CC NON_TER 1
 CC SEQUENCE 214 AA; 23230 MW; 3C0E6A85 CRC32;
 NU6M_ANSC Length: 214 February 14, 2000 08:02 Type: P Check: 1655 ..

1 GSYDLOPSLN TSYLNTWALL LTMATAPFA TYSIRMTIIV QAGOTRIPPM
 51 VSMNENPLI TAPLRLALG STAGMITS FITPKTPPM TMLITPTAA
 101 ILMITIGIIL AELSMTHT LTPKPPLM NISSALGYEN PLVHFFSKN
 151 LLEKGNIAL HLDLSWLK MGPEGLAELQ VASKAATSM HTGLIKAYLG
 201 SFALSIYMI LMTM

11AA_SEQUENCE 1.0 STANDARD; PRT; 167 AA.
 ID NU6M_APLI

AC P34857;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 6 (EC 1.6.5.3).
 GN ND6
 OS Apis mellifera ligustica (Common honeybee).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Apoidea; Apidae; Apis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=THORAX;
 RX MEDLINE: 93114603.
 RA CROZIER R.H., CROZIER Y.C.;
 RT "The mitochondrial genome of the honeybee Apis mellifera: complete
 RT sequence and genome organization";
 RL Genetics 133:97-117(1993).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L06178; AAB96808.1; -
 CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 CC KW
 CC SEQUENCE 167 AA; 19971 MW; 36B64E1B CRC32;
 NU6M_APLI Length: 167 February 14, 2000 08:02 Type: P Check: 6545 ..

1 MMLIIMLSK IFMSLSISMI LTIYLNIFN SPSMLIYLI SYSIYMSIM

51 FTWCSNMSLS IMLIIVFLS GMLIMFSYFI SLINELPLKX MKPFIOTLFL
 101 IITMKIYK LSONEHYFNY FKNIDMLVY MKNLSLFLFI MIMLIITLI
 151 LMTKIYIEK KTLRKKK

11AA_SEQUENCE 1.0 STANDARD; PRT; 263 AA.
 ID NU6M_NEUCR

AC P40915;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE 24 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
 DE (EC 1.6.99.3).
 GN NUO-24.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 OC Sordariatales; Sordariaceae; Neurospora.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA AZEVEDO J.E., DUARTE M., BELO J.A., WERNER S., VIDEIRA A.;
 RX MEDLINE: 95034970.
 RT "Complementary DNA sequences of the 24 kDa and 21 kDa subunits of
 RT complex I from Neurospora".
 RL Biochim. Biophys. Acta 1168:159-161(1994).
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE. COMPONENT OF THE FLAVOPROTEIN-SULFUR (FP)
 CC FRAGMENT OF THE ENZYME.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC -1- COFACTOR: 2FE-2S IRON-SULFUR CLUSTER N 1B.
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
 CC THIS IS A COMPONENT OF THE FLAVOPROTEIN FRACTION.
 CC -1- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
 CC MITOCHONDRIAL INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 24 KD SUBUNIT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X78083; CA54990.1; -
 CC PROSITE: PS01099; COMPLEX1_24K; 1.
 DR PFAM: PF01257; COMPLEX1_24KD; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide;
 KW Iron-sulfur.
 FT TRANSIT 1 42
 FT CHAIN 43 263
 FT METAL 143 143
 FT METAL 148 148
 FT METAL 186 186
 FT METAL 190 190
 FT METAL 190 190
 SO SEQUENCE 263 AA; 28842 MW; 0C33B3CC CRC32;

NU6M_NEUCR Length: 263 February 14, 2000 08:02 Type: P Check: 3961 ..

1 MATKILPLM RIVAVRATRL STRPSTIAPV SRACLSISAR RSDIPLMHR

51 NTPDNNPPIP FKPSADNEKY IEIITKRPP QYKKAAMPL LDGGRHGCF

101 GSISVANEVA RLLEMPMRY VEVASFYMY NRTPVGKFHY QACTTPCOL

151 GGCGSDVIYK AIKEHGIQK GETTPDGLFT FIEVEICGAC ANAPVQIIND

201 DYFEDLPET IKOYLSLAKK SVTVSKAPQ PGPSGRCOTC ENAAGLSLTL

251 SEYGPETVR SDL

11AA_SEQUENCE 1.0 STANDARD; PRT; 215 AA.
 ID NU6M_BRAOL
 AC P42027;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE 20 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
 DE (EC 1.6.99.3) (COMPLEX I-20KD) (CI-20KD).

OS Brassica oleracea (Cauliflower).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 CC Brassica.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SHOON; TISSUE-FLORETS;
 RX MEDLINE: 95334517.
 RA POGSON B.J., DOWNS C.G., DAVIES K.M., MORRIS S.C.,
 RA BUCHANAN-WOLLASTON V.,
 RT "Nucleotide sequence of a cDNA clone from broccoli with high identity
 RT with the PST subunit of NADH:ubiquinone oxidoreductase";
 RL Plant Physiol. 108:859-860(1995).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -1- COFACTOR: MAY CONTAIN AN IRON-SULFUR 4FE-4S CLUSTER.
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 20 KD SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use, by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X82274; CAA57725.1; -
 CC PROSITE: PS01150; COMPLEX1_20K; 1.
 DR PFMW: PFO1058; oxidoreductase; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transf. peptide;
 KW Iron-sulfur; 4Fe-4S.
 FT TRANSIT ? 215
 FT CHAIN 1
 FT NADH-UBIQUINONE OXIDOREDUCTASE 20 KD
 FT SUBUNIT.
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 185 185
 FT METAL 185 185
 FT SEQUENCE 215 AA; 23839 MW; DDDA728B CRC32;
 SQ
 NUKM_BROOL Length: 215 February 14, 2000 08:02 Type: P Check: 9647 ..

1 MAMTRNTAT RLPLVLOSHR AAYVSHLTS LPALSPATTP TSYTPGPPS
 51 TSAPPGISK TAEFVTSKYD DLMKARGS IMPPTTEFLAC CAVEAMHIGA
 101 ARYDLDFGI IFRSPROSD CMYAGILT N KMAPALRKVT DDMPEPRWT
 151 SMGSCANGG YHYSTSVVR GCDRIYVDI YVPCPTPAE ALLYGLLOLO
 201 KKNRKDEL HMVNR
 11AA_SEQUENCE 1.0
 ID NUOJ_ECOLI STANDARD; PRT; 184 AA.
 AC P33605; P78236;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NADH DEHYDROGENASE I CHAIN J (EC 1.6.5.3) (NADH-UBIQUINONE
 DE OXIDOREDUCTASE CHAIN 10) (NUO10).
 GN NUOJ.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / AN387;
 RX MEDLINE: 93369724.
 RA WEDNER U., GEIER S., FLOCK A., FRIEDRICH T., LEIF H., WEISS H.,
 RT "The gene locus of the proton-translocating NADH: ubiquinone
 RT oxidoreductase in Escherichia coli. Organization of the 14 genes and
 RT relationship between the derived proteins and subunits of

RT Mitochondrial complex I".
 RL J Mol. Biol. 233:109-122(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 9742617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA KILLEY M., COLLADO-VIDES J., GLASNER R.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.,
 RT "The complete genome sequence of Escherichia coli K-12".
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA ALBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ITOH T., KANAI K., KASAI H.,
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.,
 RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE. DOES COUPLE THE REDOX REACTION TO PROTON
 CC TRANSLOCATION AND THUS CONSERVES THE REDOX ENERGY IN A PROTON
 CC GRADIENT.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS. SUBUNITS NUO4, H, J,
 CC K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: TO POLYPEPTIDE 6 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
 CC OF CHLOROPLASTS OR MITOCHONDRIA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use, by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X68301; CAA48369.1; -
 CC EMBL: AE000317; AAC75340.1; -
 DR EMBL: D90859; CAB22037.1; -
 DR PIR: S37067; S37067.
 DR PIR: S38319; S38319.
 DR ECGENE: EGI2090; NUOJ.
 DR PFMW: PFO0499; oxidoreductase; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Transmembrane.
 FT TRANSMEM 1 21
 FT TRANSMEM 28 48
 FT TRANSMEM 54 74
 FT TRANSMEM 92 112
 FT TRANSMEM 138 158
 FT TRANSMEM 185 185
 FT CONFLICT 126 126
 FT CONFLICT 126 126
 FT SEQUENCE 184 AA; 19874 MW; 29FD512 CRC32;
 SQ
 NUOJ_ECOLI Length: 184 February 14, 2000 08:02 Type: P Check: 2909 ..

1 MEFAFYICGL IAILATLTVI THNPVHAL YLISLTAIS GVFFSLGAYF
 51 AGALEITIVA GAIVLFEV VMMLNLGSE IEGERQWLKP QWIGPAILS
 101 AMLVYIIVA ILGVNDGID GTPISAKAVG IILFGYVLA VELASMLLA
 151 GLVAFHVR ERAGEVLSN RKDSARAKT ESHA
 11AA_SEQUENCE 1.0
 ID NUOJ_ECOLI STANDARD; PRT; 201 AA.
 AC P19968;
 DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MADH-UBIQUINONE OXIDOREDUCTASE 21.3 KD SUBUNIT (EC 1.6.5.3)
 DE (EC 1.6.99.3).
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 CC Sordariaceae; Neurospora.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 90147712.
 RA VIDEIRA A., TROPSCHUG M., WERNER S.;
 RT "Primary structure, in vitro expression and import into mitochondria
 RT of a 29/21-kDa subunit of complex I from Neurospora crassa.";
 RL Biochem. Biophys. Res. Commun. 166:280-285(1990).
 CC -1- FUNCTION: TRANSFER OF ELECTRON FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC -1- SUBUNIT COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE. IMPORTATION
 CC INTO MITOCHONDRION IS MEMBRANE POTENTIAL DEPENDENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M32244; AAA33570.1; -
 CC PIR: A34051; A34051.
 CC DR OXIDOREDUCTASE; NAD; UBIQUINONE; Mitochondrion; Membrane.
 CC KW SEQUENCE 201 AA; 21347 MW; 32F6D36F CRC32;
 SQ
 NUZM_NEUCR Length: 201 February 14, 2000 08:02 Type: P Check: 7750 ..
 1 MASKVTVGV KTAGGVV PV SQKTVQSG VMERIRRAFA IDPKRSGV
 51 LVYNNRNPSP GSLDPLAYDD PVTPPADIA DNPYKRRAR RNYRRLSYGV
 101 QAQAVALLSY GSATHPRVEL VGENGSKOLV AAOENKGTGG LAKTFESTGV
 151 EAKVLVALET GGLPPLPESG KLGEGKWDV YVYQIAEELS YSEAYPCRSF
 201 S
 !!AA_SEQUENCE 1.0 STANDARD: PRT; 306 AA.
 ID OPB_ECOLI
 AC P31132; P76026; P77550;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE OLIGOPETIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPB.
 GN OPB.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,

RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MICROBUCCHI K., MORI H., MOTOKURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SATO N., SAMPEI G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-16 FROM N.A.
 RA MEDLINE: 90256748.
 RA KASHIMOTO K., TAGUCHI Y., SAKAI Y., KOBAYASHI H., IGARASHI K.;
 RT Identification of the polyamine-induced protein as a periplasmic
 RT oligopeptide binding protein.";
 RL J. Biol. Chem. 265:8387-8391(1990).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR OLIGOPETIDES; PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
 CC THE SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPBPC
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000223; AAC74326.1; -
 CC EMBL: D90763; CAB20709.1; -
 CC DR EMBL: D90852; CAB21967.1; -
 CC DR EMBL: J05433; -; NOT_ANNOTATED_CDS.
 CC PIR: B36263; B36263.
 CC DR ECOCENF; PG10675; OPB.
 CC DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER: 1.
 CC DR PIRAF; PF00528; BPD_TRANSP: 1.
 CC KW Transport; Peptide transport; Transmembrane; Inner membrane.
 CC FM DOMAIN 1
 CC FT TRANSMEM 1 9
 CC FT DOMAIN 1 9
 CC FT TRANSMEM 10 30
 CC FT DOMAIN 31 99
 CC FT TRANSMEM 100 121
 CC FT DOMAIN 122 137
 CC FT TRANSMEM 138 158
 CC FT DOMAIN 159 172
 CC FT TRANSMEM 173 190
 CC FT DOMAIN 191 226
 CC FT TRANSMEM 227 250
 CC FT DOMAIN 251 271
 CC FT TRANSMEM 272 293
 CC FT DOMAIN 294 306
 CC SQ SEQUENCE 306 AA; 33443 MW; 3FCC09F5 CRC32;
 OPB_ECOLI Length: 306 February 14, 2000 08:02 Type: P Check: 5147 ..
 1 MKKFLRCL EALPILFILI TISFFMRLA PGSPFPGENT LPPEVMANIE
 51 AKYHLNPPIM TQYFSLKOL AHGDFGSEK YKDYSDNDLV ASSPVSAAKL
 101 GAAPFLAVI LGVSAGVIA LKONTKWDT VAGLANTGV IPSTVAPPL
 151 VMFAILLHW LPEGGNGGA LKEMILPWA LSLAYASTIA RITGSMIEV
 201 LHSNFIART AKGLPRLII LRHALKPLL PVLSTGPAF VGIITGSMVI
 251 ETIGLPGIG QLVNGALNR DYSLVLSLI LVGALTIIFN AIVDVLAVI
 301 DPKIRY
 !!AA_SEQUENCE 1.0 STANDARD: PRT; 306 AA.
 ID OPB_SALT
 AC P08005;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)

01-FEB-1995 (Rel. 31, last annotation update)
DE OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPB.
GN OPB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN:LT2
RX MEDLINE: 86011222
RA HIGGS I.D., GALLAGHER M.P., JAMESON D.J., HIGGINS C.F.;
RT "Molecular characterization of the oligopeptide permease of
Salmonella typhimurium.";
RL J. Mol. Biol. 195;125-142(1987).
RN (2)
RP TOPOLOGY.
RX MEDLINE: 92149312.
RA PEARCE S.R., MIMACK M.L., GALLAGHER M.P., GLEADI U., HYDE S.C.,
HIGGINS C.F.;
RT "Membrane topology of the integral membrane components, OPB and
OppC, of the oligopeptide permease of Salmonella typhimurium.";
RL Mol. Microbiol. 6:47-57(1992).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
FOR OLIGOPEPTIDES. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPBPC
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05491; CAA29040.1; -
DR PIR: B29333; OREBOB.
DR SYNGENE: SG10268; OPB.
DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; 1.
DR PFAM: PF00528; BPD_transp; 1.
KW Transport; Peptide transport; Transmembrane; Inner membrane.
FT DOMAIN 1 9
FT TRANSMEM 31 99
FT DOMAIN 100 121
FT TRANSMEM 122 137
FT TRANSMEM 138 158
FT TRANSMEM 159 172
FT TRANSMEM 173 190
FT TRANSMEM 191 226
FT TRANSMEM 227 250
FT DOMAIN 251 271
FT TRANSMEM 272 293
FT TRANSMEM 294 306
FT SEQUENCE 306 AA; 33420 MW; 8A020E68 CRC32;
OPB_SALTY Length: 306 February 14, 2000 08:02 Type: P Check: 4920 ..

1 MKFILLRCL EAIPILFILI TISFFMRRLA PGSPFTGERA LPPEVLNIE
51 AKYHLNDPIM TOYFSYKOL AHGDFGSPK YKDYVNDV AASEPVSATL
101 GAFAFLAVI IGVSAVIA LKONTWMDY VMGFATGVV IPSEFVAPL
151 VMVAFLIOW LPGGWNNGA LKFMILPVA LSLAYASIA RITRSMIEV
201 LHSFIFITAR AKGLPWRRII FRHALKRAL PVLSTYGPV VGIIGSMVI
251 ETIIGLPGIG QLFVNGALNR DYSVLVSLTI LVGALLILEN AIVDVLAVI

301 DEKIRY
!!AA_SEQUENCE 1.0
ID OPB1_CALVI STANDARD; PRT; 371 AA.
AC P22269;
DT 01-AUG-1991 (Rel. 19, Created)
DR 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE OPSTN RH1 (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
GN RH1 OR NINAE
OS Calliphora vicina (Blue blowfly) (Calliphora erythrocephala).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Fterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Coleoptera; Calliphoridae; Calliphora.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 91009258.
RA HUBER A., SMITH D.P., ZUKER C.S., PAULSEN R.;
RT "Opsin of Calliphora peripheral photoreceptors R1-6. Homology with
Drosophila R1 and posttranslational processing.";
RL J. Biol. Chem. 265:17906-17910(1990).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
BE PHOSPHORYLATED.
CC -1- MISCELLANEOUS: EACH EYE IS COMPOSED OF 800 FACETS OR OMMATIDIA.
CC R6 CELLS ARE OUTER CELLS. WHILE R7 AND R8 ARE INNER CELLS.
CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58334; AAA62725.1; -
DR PIR: A39234; A39234.
DR GCRDB: GCR_0009; -
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR; 1.
DR PROSITE: PS00238; OPSIN; 1.
DR PFAM: PF00001; 7tm.1; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
KW Glycoprotein; G-protein coupled receptor; Vision.
FT DOMAIN 1 48
FT TRANSMEM 49 72
FT DOMAIN 73 84
FT TRANSMEM 85 110
FT DOMAIN 111 124
FT TRANSMEM 125 144
FT DOMAIN 145 163
FT TRANSMEM 164 187
FT DOMAIN 188 211
FT TRANSMEM 212 239
FT TRANSMEM 240 274
FT DOMAIN 275 298
FT TRANSMEM 299 305
FT TRANSMEM 306 330
FT TRANSMEM 331 371
FT CARBOHYD 18 18
FT DISULFID 121 18
FT BINDING 317 317
FT SEQUENCE 371 AA; 41289 MW; 6D98882F CRC32;
OPB1_CALVI Length: 371 February 14, 2000 08:02 Type: P Check: 9791 ..

1 MERYSTPLIG PSFALTNNGS VTDKVTDPMA HLHVRYWNOF PAMEPRWAKF

51 LAAYWLVIAT ISMCGNGVY YIFSTTKSLR TPANLVINL AISDFGIMIT
 101 NTPMGINLF YETWVGLPLM CDYIGGLGSA FGCSSILSMC MISLDRYNVI
 151 VGMAGQPMY IKLAIMKIAL IMFMASIMTL APVGSMSRYV PEGNLTSCGI
 201 DYLERDMNRR SYLIFYSIFV YLPLFLICY SWFLLAASV AHKARBEQA
 251 KKMANKSLRS SEDADKSAEG KLAKVALYTI SLMFAMTPY TIINTGLFR
 301 YGLPLPLNTI WGACFAKSA CYNPIYGIS HPKYGIALKE KCPCCVFGVY
 351 DDGKASDATS QATNNESETK A

11AA_SEQUENCE 1.0
 ID OPS1_DROME STANDARD: PRT: 373 AA.
 AC P06002: 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE OPSIN RH1 (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
 GN RH1 OR NINAE.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85176937.
 RA O'LOUSA J.E., BAEHR W., MARTIN R.L., HIRSH J., PAK W.L.,
 RA APPLEBURY M.L.;
 RT "The Drosophila rh1 gene encodes an opsin.";
 RL Cell 40:839-850(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85176938.
 RA ZUKER C.S., COMAN A.F., RUBIN G.M.;
 RT "Isolation and structure of a rhodopsin gene from D. melanogaster.";
 RL Cell 40:851-858(1985).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- MISCELLANEOUS: EACH DROSOPHILA EYE IS COMPOSED OF 800 FACETS OR
 CC OMATIDIA. EACH OMATIDIUM CONTAINS 8 PHOTORECEPTOR CELLS (R1-R8),
 CC THE R1 TO R6 CELLS ARE OUTER CELLS, WHILE R7 AND R8 ARE INNER
 CC CELLS.
 CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: K02315; AAA28733.1; -;
 CC EMBL: K02320; AAA28735.1; ALT-SEQ.
 CC EMBL: K02316; AAA28735.1; JOINED.
 CC EMBL: K02317; AAA28735.1; JOINED.
 CC EMBL: K02318; AAA28735.1; JOINED.
 CC EMBL: K02319; AAA28735.1; JOINED.
 CC PIR: A22012; COFE.
 CC GCRDB: GCR 0028; -;
 CC GCRDB: GCR 0029; -;
 CC FLVBASE: FBgn0002940; rh1ae.
 CC PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.

DR PROSITE: PS00238; OPSIN; 1.
 KW PFM; PFO0001; 7tm_1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
 KW Glycoprotein; G-protein coupled receptor; Vision.
 FT DOMAIN 1 49
 FT TRANSMEM 50 74
 FT DOMAIN 75 86
 FT TRANSMEM 87 112
 FT DOMAIN 113 126
 FT TRANSMEM 127 146
 FT DOMAIN 147 165
 FT TRANSMEM 166 189
 FT DOMAIN 190 213
 FT TRANSMEM 214 241
 FT DOMAIN 242 276
 FT TRANSMEM 277 300
 FT DOMAIN 301 307
 FT TRANSMEM 308 332
 FT DOMAIN 333 373
 FT CARBOHYD 20 20
 FT DISULFID 123 200
 FT BINDING 319 319
 SO SEQUENCE 373 AA; 41494 MW; 43B10BFA CRC32;

OPS1_DROME Length: 373 February 14, 2000 08:02 Type: P Check: 4358 ..

1 MESFAVAQAQ LGPHFAPLSN GSVVDKVTPT MAHLSPYWN QEPAMDPIWA

51 KILFAYIMI GMSMCGNV YIFATTKS LTPANLVNI NLATSDFGIM

101 ITNTPMGIN LFETWVGLP MMCDIYAGLG SAFGCSSTWS MGMTSLDRYO

151 VIVMGAGRP KTIPLALSKI AYIMFSSIM CLAPFGNSR YVPGNLTSC

201 GIDLERDMN PRSYLIFSI FVYIPLFLI CYSYFIINA VSAHEKAMRE

251 QAKMANKSL RSEDEAKSA EKLAKVALY TITLWMAMT PYLVINCMGL

301 FREFGLPLN TIWGAFAKS AACNPIYVG ISHPRYRLAL KKPCCVFGV

351 KVDDKSSDA QOATASENAE KSA

11AA_SEQUENCE 1.0
 ID OPS1_DROME STANDARD: PRT: 374 AA.
 AC P28678;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE OPSIN RH1 (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
 GN RH1 OR NINAE.
 OS Drosophila pseudoobscura (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93012921.
 RA CARULI J.P., HARTL D.L.;
 RT "Variable rates of evolution among Drosophila opsin genes.";
 RL Genetics 133:193-204(1992).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- MISCELLANEOUS: EACH DROSOPHILA EYE IS COMPOSED OF 800 FACETS OR
 CC OMATIDIA. EACH OMATIDIUM CONTAINS 8 PHOTORECEPTOR CELLS (R1-R8),
 CC THE R1 TO R6 CELLS ARE OUTER CELLS, WHILE R7 AND R8 ARE INNER
 CC CELLS.
 CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

OPSIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X65877; CAA46708.1; -
 CC PIR: S24605; S24605.
 CC GCRDB: GCR_0509; -
 CC GCRDB: GCR_0429; -
 CC DR FLYBASE: FBgn0012733; Dpse.unnae.
 CC DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
 CC DR PFM: PFO0001; 7tm_1; 1.
 CC DR PFM: PFO0001; 7tm_1; 1.
 CC KW Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
 KW Glycoprotein; G-protein coupled receptor; Vision.
 FT DOMAIN 1 49 EXTRACELLULAR
 FT TRANSMEM 50 74 CYTOPLASMIC
 FT TRANSMEM 75 86 2 (POTENTIAL)
 FT TRANSMEM 87 112 2 (POTENTIAL)
 FT TRANSMEM 113 126 EXTRACELLULAR
 FT TRANSMEM 127 146 3 (POTENTIAL)
 FT TRANSMEM 147 165 CYTOPLASMIC
 FT TRANSMEM 166 190 4 (POTENTIAL)
 FT TRANSMEM 191 214 EXTRACELLULAR
 FT TRANSMEM 215 242 5 (POTENTIAL)
 FT TRANSMEM 243 277 CYTOPLASMIC
 FT TRANSMEM 278 301 6 (POTENTIAL)
 FT TRANSMEM 302 308 EXTRACELLULAR
 FT TRANSMEM 309 333 7 (POTENTIAL)
 FT TRANSMEM 334 374 CYTOPLASMIC
 FT CARBOHYD 20 20 PROBABLE
 FT DISULFID 123 201 POTENTIAL
 FT BINDING 320 320 RETINAL CHROMOPHORE
 SQ SEQUENCE 374 AA; 41604 MW; AAC8CB5 CRC32;

OSL_Drops Length: 374 February 14, 2000 08:02 Type: P Check: 7241 ..

1 MOSFAVATQ LGPOFAFNS GSVDKVTDP MAHLISPYMD QPAMPPIVA
 51 KILTAVMIII GMISWNGV VYIPATKS LRTANLLVI NLAISDFGIN
 101 INTFMGNG LNFTWTWLG MMDIYAGIG SAFGCSSTWS MCMISIDRQ
 151 VYKMGARP MTPLALGKI AYIMESTIM CCLAPYVGS RYVPEGLUS
 201 CGIDLERDW NPSYLIFFS IYVYIPLF ICYSWFLIA AVAHEKAR
 251 EOKAKMNVKS LRSESDAKS AEGKLAKVAL VTISLFWAM TPLVINCWG
 301 LKFEGLTPL NTIWACFAK SAACYNPIVY GISHPKRYLA LKEKPCCVF
 351 GKYDDKSSE AQSQTSEA ESKA
 !!AA_SEQUENCE 1.0
 ID OTCA_VIBS2 STANDARD; PRT; 301 AA.
 AC P96172;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE ORNITHINE CARBAMOYLTRANSFERASE, ANABOLIC (EC 2.1.3.3) (OTCASE).
 GN ARGF.
 OS Vibrio sp. (strain 2693).
 CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibri.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LIANG Z.Y., DEMAREZ M., LECRAIN C., BAETENS M., ROOYERS M.,
 RA GLASDORF N.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -1 CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1 PATHWAY: ARGININE BIOSYNTHESIS.
 CC -1 SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1 SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Y11033; CA41921.1; -
 CC DR HSSP: P04391; 1AKM.
 CC DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
 CC DR PFM: PFO0185; OTCase; 1.
 CC KW Arginine biosynthesis; Transferase.
 SQ SEQUENCE 301 AA; 32825 MW; 3CA3E90D CRC32;

OTCA_VIBS2 Length: 301 February 14, 2000 08:02 Type: P Check: 7498 ..

1 MENLSYKDL SKQIIDLIA LAKAVKANPA EYSQALAGS IYIYERSL
 51 RRTYTPDIGI HKLGGHAYL DAONGAIGER EYKDFRANI SSMADAIYAR
 101 VYSHKTEGL VEGSYPPYV SLCDLYHPCQ ALADFLISE HYEDVSKYKL
 151 AYVEGNNVT HSLMTGAIL GAEYVACPR GSSPDQIYK QAMALAEISG
 201 KRINTDNL DAIYDYVITG DTWWSGGDT PLAQVEKEM PYQINKALLM
 251 RTGIKHLVLC QPAHRELEIT SEVMDGHSIL IEDQENRHH AONAVLLTL
 301 K

!!AA_SEQUENCE 1.0
 ID OTCA_SCHPO STANDARD; PRT; 327 AA.
 AC P31317;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE PRECURSOR (EC 2.1.3.3) (OTCASE)
 GN (ORNITHINE TRANS-CARBAMYLASE).
 OS Schizosaccharomyces pombe (fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Archiaascmycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-975;
 RC MEDLINE: 92209520.
 RA VAN HOFFEL C., DOBOIS E., MESSENGUY F.;
 RT Cloning and sequencing of arg3 and arg11 genes of
 RT Schizosaccharomyces pombe on a 10-kb DNA fragment. Heterologous
 RT expression and mitochondrial targeting of their translation
 RT products.
 RL Eur. J. Biochem. 205:33-43(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-972;
 RC CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -1 CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1 PATHWAY: FIRST STEP IN ARGININE BIOSYNTHESIS.
 CC -1 SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1 SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC The Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63577; CAA45133.1; -
DR EMBL: X63727; CAA93560.1; -
DR PIR: S2290; OW2P.
DR HSP: Q51742; IAL5.
DR PROSITE: PS00097; CARBAMOYLTRANSFERASE: 1.
DR PRAM: PF00185; OTCace: 1.
KW Arginine biosynthesis; Transferase; Mitochondrion; Transit peptide.
FT TRANSIT 1 MITOCHONDRION.
FT CHAIN 1 327 ORNITHINE CARBAMOYLTRANSFERASE.
SQ SEQUENCE 327 AA; 36119 MW; 9CC43E5C CRC32;
OTC_SCHPO Length: 327 February 14, 2000 08:02 Type: P Check: 9240 ..
1 MSFKFPRHL LSIKIDSRGE IVKIDRSSE IKOAYKONFQ NRSYOMSGL
51 SSCWAMIFS KRSTRRSV SAVSCLGN AMFLKDDIQ LGVNESTLYDT
101 SKVSSWVSG IYAVNKYSD VATLAKHASC PVINGLDDTF HPLQALADLL
151 TIKETFSFD GLKVAWVGA NNVLHDLMLA NAKVGIHTSV AKPKOVNVD
201 DILSTVNEA NENGSTFEIV NDPVAVKNA DIVVDTWIS MGOEAKRQR
251 LKOTGQVYV GEIMKLAKPS CKFMHCLPRH PEEVSDEVY GENSLVPEEA
301 ENKRTIVAV LEALVNRGE ILPPASA
11AA_SEQUENCE 1.0
ID OXDD_BOVIN STANDARD; PRT; 341 AA.
AC P31228; 002846; -
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE D-ASPARTATE OXIDASE (EC 1.4.3.1) (DASOX) (DDO).
GN DDO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE OF 1-338.
RC TISSUE-KIDNEY;
RX MEDLINE: 92291057.
RA NEGRI A., CECILIANI F., TEDESCHI G., SIMONIC T., RONCHI S.;
RT "The primary structure of the flavoprotein D-aspartate oxidase from
RT beef kidney".
RT J. Biol. Chem. 267:11865-11871(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY CORTEX;
RX MEDLINE: 97220379.
RA SIMONIC T., DUGA S., NEGRI A., TEDESCHI G., MALCOVANTI M.,
RA TEICHRINI M.L., RONCHI S.;
RT cDNA cloning and expression of the flavoprotein D-aspartate oxidase
RT from bovine kidney cortex.
RL Biochem. J. 322:729-735(1997).
CC -1- CATALYTIC ACTIVITY: D-ASPARTATE + H(2)O + O(2) = OXALACETATE +
CC NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD OR 6-HYDROXYFLAVIN ADENINE DINUCLEOTIDE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X85310; CAA64622.1; -
DR PIR: A4432; A4432.
DR HSP: P00371; IDAO.
DR PROSITE: PS00677; DAO: 1.
DR PRAM: PF00342; MICROBODIES_CTER: 1.
DR PRAM: PF01265; DAO: 1.
KW Oxidoreductase; Flavoprotein; FAD; Peroxisome.
FT MOD_RES 1
FT NE_BIND 6 20
FT ACT_SITE 223 223 FAD (ADP PART) (POTENTIAL).
FT ACT_SITE 302 302 BY SIMILARITY.
FT SITE 339 341 MICROBODY TARGETING SIGNAL (POTENTIAL).
FT VARIANT 228 228 V -> I (IN SOME MOLECULES).
FT CONFLICT 274 274 K -> R (IN REF. 1).
FT CONFLICT 283 283 S -> G (IN REF. 1).
SQ SEQUENCE 341 AA; 37659 MW; 842B960C CRC32;
OXDD_BOVIN Length: 341 February 14, 2000 08:02 Type: P Check: 9943 ..
1 MDVRLAVVG AGVGLSTAV CISKWPGCS ITVISDKFTP ETTSDVAAGM
51 LIPTYPDTP IQKQKWEKE TFDHLEAIYN SAEADAGVI LVSGMQLFQS
101 IPTPEVPYNA DVLGFRKMT KDELKFPQH VEGHATFLK CEGPAVILPWL
151 QKRVKNGSL ILTRIDELW ELHPSFDIV NCSGLSGHQL AGSKTFPVR
201 GQVLKQAPW VKHTRDSSG LTIYVGVSN VTLGSTRKG DWNISPDAEI
251 SKELISRCCA LPSLRGAYD LREKVLRLPT RPSVLEKEL LAODSRRLPV
301 VHHYGHSGG IAHMNGTAL ATRLVNECQV VLRTPAKSK L
11AA_SEQUENCE 1.0
ID OXDD_HUMAN STANDARD; PRT; 341 AA.
AC Q99489;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE D-ASPARTATE OXIDASE (EC 1.4.3.1) (DASOX) (DDO).
GN DDO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97306065.
RA SETOVANA C., MIURA R.;
RT "Structural and functional characterization of the human brain D-
RT aspartate oxidase".
RT J. Biochem. 121:798-803(1997).
CC -1- FUNCTION: SELECTIVELY CATALYZES THE OXIDATIVE DEAMINATION OF D-
CC ASPARTATE AND ITS N-METHYLATED DERIVATIVE, N-METHYL D-ASPARTATE.
CC -1- CATALYTIC ACTIVITY: D-ASPARTATE + H(2)O + O(2) = OXALACETATE +
CC NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD OR 6-HYDROXYFLAVIN ADENINE DINUCLEOTIDE.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS, DDO-1 (SHOWN HERE) AND DDO-2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@lsb-sib.ch.

CC -----
 CC EMBL: D89858; BA14031.1;
 CC HSSP: P00371; IDAO.
 DR MIM: 124450;
 DR PROSITE: PS00677; DAO: 1.
 DR PFAM: PF01266; DAO: 1.
 KW Oxidoreductase; Flavo-protein; FAD; Peroxisome; Alternative splicing.
 FT NP_BIND 6 20
 FT ACT_SITE 223 223 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT SITE 338 341 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT SITE 35 153 MISSING (IN ISOFORM DDO-2).
 FT SEQUENCE 341 AA; 37535 MW; 0948D6AB CRC32;
 SQ
 OXND_HUMAN Length: 341 February 14, 2000 08:02 Type: P Check: 7167 ..

1 MOTARLAAYV AGVGLSTAV CISKIVRCS VTIIDKFTP DTISDVAGM
 51 LIPHTYDTP IHTOKMRE TENHPIRAN SAEADAGVH LVSGWQIFGS
 101 TPTVEVPEVA DVLGFRMT EAIKKEPQY VEGQAFILK CECPAIPML
 151 EKIKSGGW TLTRIEDLW ELHPSDIW NCSGLSROL AGDSKIPVR
 201 GOVLQVQAPV VEHFRDGS LTYIPEGTSH VILGTRQKG DMNLSPAEN
 251 SREILSRCA LEPSLHACN IREKVGRLPY RGVRLQTEL LARDGORLPV
 301 VHHYHSGSG ISVHWGTALE AARLVSECVH ALRPIPKSN L

11AA_SEQUENCE 1.0
 ID P2Y6_HUMAN STANDARD: PRT; 328 AA.
 AC 015077; 015754;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE P2Y PURINOCEPTOR 6 (P2Y6).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CN [1]
 RA SEQUENCE FROM N.A.
 RA COMMUNI D., PARMENTIER M., BOEYNAENS J.M.;
 RA Submitted (MAY-1996) to the EMBL/Genbank/DBD databases.
 RN [2]
 RX SEQUENCE FROM N.A.
 RX MEDLINE: 97432828.
 RA SOMERS G.R., HAMMET F., WOOLLIATT E., RICHARDS R.I., SOUTHEY M.C.,
 RA VENTER D.J.;
 RA "Chromosomal localization of the human P2y6 purinoceptor gene and
 RA phylogenetic analysis of the P2y purinoceptor family";
 RA Genomics 44:127-130(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP TISSUE-PLACENTA;
 RC MEDLINE: 97415792.
 RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 RA "Cloning of P2y6 cDNAs and identification of a pseudogene: comparison
 RA of P2y receptor subtype expression in bone and brain tissues.";
 RA Biochem. Biophys. Res. Commun. 237:297-302(1997).
 RN [4]
 RP ERRTUM.
 RP MEDLINE: 98069816.
 RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 RA Biochem. Biophys. Res. Commun. 240:298-302(1997).
 CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
 CC PHOSPHATIDYLINOSITOL-CALCIN SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

CC -----
 CC EMBL: X97058; CA65770.1;
 CC EMBL: U52464; AAB03572.1;
 CC EMBL: AF007892; AAB0713.1;
 CC EMBL: AF007891; AAB0712.1;
 CC HSSP: P34996; IDDO.
 DR MIM: 602451;
 DR PROSITE: PS00237; G-PROTEIN-RECEPTOR; FALSE-NEG.
 DR PFAM: PF00001; 7tm.1; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 27
 FT TRANSSEM 28 48
 FT DOMAIN 49 62
 FT TRANSSEM 63 83
 FT DOMAIN 84 101
 FT TRANSSEM 102 122
 FT DOMAIN 123 144
 FT TRANSSEM 145 165
 FT DOMAIN 166 194
 FT TRANSSEM 195 215
 FT DOMAIN 216 236
 FT TRANSSEM 237 257
 FT DOMAIN 258 280
 FT TRANSSEM 281 303
 FT DOMAIN 304 328
 FT DISULFID 99 177
 FT CARBOHYD 2 5
 FT CONFLICT 5 3
 FT SEQUENCE 328 AA; 36429 MW; 29C125CE CRC32;
 SQ
 P2Y6_HUMAN Length: 328 February 14, 2000 08:02 Type: P Check: 7545 ..

1 MEMDNGTGOA LGIPPTTCY RENEKOLLP PYVSAYVANG LPLNCIVTO
 51 ICTSRALTR TAVYTTLNLAL ADLYACSLP LLIYNAQSD HMPFEDFACR
 101 LVRFLEFVNL HGSILFLTCI SFORYLGICH PLAPWHRKG RRAANLVCA
 151 VWLAATTOCL PLAIFATGI QNRRTVCYDL SPALATHYV PYGALTYIG
 201 FLPPFAALLA CYCLACRLC RODGPAEVA QERGRKARV AVVAAPAI
 251 SLPFHIRT AYLAVERSTPG VPCVLEAFA AAYKGRPFA SANVYLDPL
 301 FYTQKFRF RPHLLQKLT AKWROGR

11AA_SEQUENCE 1.0
 ID P2Y6_HUMAN STANDARD: PRT; 166 AA.
 AC P31078;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE P2Y PURINOCEPTOR 6 (P2Y6).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CN [1]
 RA SEQUENCE FROM N.A.
 RA STRAIN-MT1131;
 RC MEDLINE: 92356828.
 RA TOKITO M.K., DALDAL F.;
 RA "petr, located upstream of the fbcrC operon encoding the cytochrome
 RA bcl complex, is homologous to bacterial response regulators and
 RA necessary for photosynthetic and respiratory growth of Rhodospirillum rubrum
 RA capsulatus";

```

RL MOL. Microbiol. 6:1645-1654(1992).
CC -1- FUNCTION: NECESSARY FOR PHOTOSYNTHETIC AND RESPIRATORY GROWTH.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z12113; CAAT8097.1; -.
DR PIR: S22631; S22631.
DR PROSITE: PS01117; HTH_MARR_FAMILY; 1.
DR PFAM: PF01047; Marr; 1.
KW Transcription regulation; DNA-binding.
SO SEQUENCE 166 AA; 18807 MW; 99E3E86E CRC32;

PEPT_RHOC Length: 166 February 14, 2000 08:02 Type: P Check: 1875 ..

1 MADTGAPGGE TLFLTDEQL RKGIEAMFPA YRGFTADPDR ILDOHDYGRA
51 HRRATHTNR EPGLTVTLLI SYLGVTKOSI NRVLTLLID GLVESRYGR
101 DRRERHLHT EKGAVEREL SEQRVMPA AYRAQPAV AGFQVLEAM
151 MDPARRHYQ MLKDAE

11AA SEQUENCE 1.0 STANDARD: PRT: 441 AA.
ID PEK3_YEAST
AC P28795;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PEROXISOMAL MEMBRANE PROTEIN PAB3 (PEROXIN-3).
GN PEK3 OR PAB3 OR IDK329C OR D9798.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
NC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91373453.
RA HOEHELD J., VEENHUIS M., KUNAU W.-H.;
RT "PAB3, a Saccharomyces cerevisiae gene encoding a peroxisomal
RT integral membrane protein essential for peroxisome biogenesis.";
RL J. Cell Biol. 114:1167-1178(1991).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN: S288C / AB972;
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAYELLO A., HULTON L., GATUNG S., GRECO T., KIRSTEN J.,
RA KUCHBA T., HALLSWORTH K., HARKINS J., HILLER L., JIR R.,
RA JOHNSON D., JOHNSON L., LANGSTON T., LATREILLE P., LE I.,
RA MARIIS E., MENEZES S., MILLER N., NHAU M., PADLEY A., PELUSO D.,
RA RIKEN L., RILES L., TAICH A., TREVASIS E., VIGNATI D.,
RA WITCOX L., WOLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. MAY FUNCTION
CC AS A RECEPTOR PROTEIN. LACK OF THIS PROTEIN CAUSES THE
CC PEROXISOMAL-DEFICIENT PHENOTYPE AND MISLOCALIZATION IN THE
CC CYTOSOL OF PEROXISOMAL MATRIX PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- INDUCTION: BY OLEIC ACID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: X58407; CA41309.1; -.
DR EMBL: U32517; AAB64764.1; -.
DR PIR: A40550; A40550.
DR SGD: L0001339; PAB3.
KW Transmembrane; Peroxisome.
FT DOMAIN 1 17 MATRIX (POTENTIAL).
FT TRANSMEM 18 39 POTENTIAL.
FT DOMAIN 40 441 CYTOPLASMIC (POTENTIAL).
SO SEQUENCE 441 AA; 50675 MW; 96C48B96 CRC32;

PEK3_YEAST Length: 441 February 14, 2000 08:02 Type: P Check: 166 ..

1 MAPNQRSRL LQRRGKGLI SLTGIALFT TGSVVFVFK RMLKQQLRI
51 TEQHFKEQI KRREQIQED SLTIYELP VQKAVLNEND LNDISIYQI
101 KQOKNQLTRA KSSSRSSSP LSKAKLWNE LELKSLILV TVYTVSSLI
151 LTRIQNLTL TRNEYLDSAI KLTMQOENR KLQNFYMWV TSWSDPEDK
201 ADDAWYMAK KSKKEQGEV INEAFSLIS WMLKGLS YNEIITQIE
251 IEFDIGHPD TLTEEFSSR LTNIFNTNS QIFQONNNL TSILPKDSS
301 GQEFLLSQTL DADALTSFHS NTLVFNQLVN ELTQIEESTA TSIVLESILN
351 ESFHFIMNKV GIKTIKKRP GQEDQOQOYM AVFAMSKDC COEMLQTTAG
401 SSHSGSVNEY LATLDSVOPL DDLASVYSN FGVSSEFSK P

11AA SEQUENCE 1.0 STANDARD: PRT: 246 AA.
ID PHB3_ALICEU
AC P14697;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ACETOACETYL-COA REDUCTASE (Ec 1.1.1.36).
GN PHB3.
OS Alcaligenes eutrophus.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
NC [1]
RP SEQUENCE FROM N.A.
RX STRAIN: H16.
RX MEDLINE: 89359356.
RA PEOPLES O.P., SINSKEY A.J.;
RT "Poly-beta-hydroxybutyrate biosynthesis in Alcaligenes eutrophus H16.
RT Characterization of the genes encoding beta-ketothiolase and
RT acetoacetyl-CoA reductase.";
RL J. Biol. Chem. 264:15293-15297(1989).
CC -1- CATALYTIC ACTIVITY: (R)-3-HYDROXYACYL-CoA + NADP(+) -> 3-OXOACYL-
CC CoA + NADPH
CC -1- PATHWAY: SECOND STEP IN POLY-BETA-HYDROXYBUTYRATE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04987; AAA21973.1; -.
DR PIR: B34340; BDALAE.
DR HSSP: P19992; 2HSD.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short_C2; 1.
DR PFAM: PF00678; adh_short_C2; 1.
KW Oxidoreductase; NADP; PHB biosynthesis.

```

FT NP_BIND 8 32 NADP (BY SIMILARITY).
 FT ACT_SITE 153 BY SIMILARITY.
 SQ SEQUENCE 246 AA: 26370 MW; B1A08AA3 CRC32;
 PHBB_ALCEU Length: 246 February 14, 2000 08:02 Type: P Check: 2829 ..

1 MGRIVYVVG GMGGIGTAIC QRLAKDFRY VAGGCPNSPR REKMLEQOKA
 51 LGDFIASGCV NVADWDSTKT APDKVSEVG EVDVLINAG ITRDYVERKM
 101 TRADMAVID TNLSTLEFNT KQYIDGADR GWRIVNISS VNGKQGFQ
 151 TNSSTAKGL HGFTMALAE VAKGVYVNI VSPGIATDM VKAIRQVLD
 201 KIVATIPYKR LGPEELASI CAVLSESEG FSTGADPSLN GGLHMG

11AA_SEQUENCE 1.0 STANDARD: PRT: 252 AA.
 ID PHNK_ECOLI
 AC 016678;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNK.
 GN PHNK.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 91193228.
 RA MAKINO K., KIM S.K., SHINAGAWA H., AMEMURA M., NAKATA A.;
 RT "Molecular analysis of the cryptic and functional *phn* operons for
 RT phosphate use in *Escherichia coli* K-12.";
 RL J. Bacteriol. 173:2665-2672(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12; MG1655;
 RX MEDLINE: 95334362.
 RA BURLAND V.D., PLUNKERT G. III, SOPHA H.J., DANIELS D.L.,
 RA BLATTNER F.R.;
 RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B;
 RX MEDLINE: 90170953.
 RA CHEN C.-M., YE Q.-Z., ZHU Z., WANNER B.L., WALSH C.T.;
 RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
 RT sequencing of the *phn* (*psid*) genes involved in alkyldiphosphate
 RT uptake and C-P lyase activity in *Escherichia coli* B.";
 RL J. Biol. Chem. 265:4461-4471(1990).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR ALKYLPHOSPHONATES; PROBABLY RESPONSIBLE FOR ENERGY COUPLING
 CC TO THE TRANSPORT SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D90227; AAA14271.1;
 DR EMBL: U16003; AAA86986.1;
 DR EMBL: A5000482; AAC77058.1;
 DR EMBL: J05260; AAA24349.1;
 DR PIR: C35719; C35719.
 DR ECOGENE: EG10720; PHNK.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR FRAM: PF00003; ABC_tran; 1.
 KW Alkyldiphosphate uptake; Transport; ATP-binding.
 FT NP_BIND 38 45
 FT CONFLICT 47 45 L -> O (IN REF. 2).
 SQ SEQUENCE 252 AA: 27831 MW; CB59EB3 CRC32;
 PHNK_ECOLI Length: 252 February 14, 2000 08:02 Type: P Check: 804 ..

1 MNOPLSYNN LTHLYAPKGG FSDVSFDLMP GEVLGIAGES GSGKTLILKS
 51 ISARLTPQOG EIHYENRSLY AMSEADRRRL LRTMGVYHO HPLDGLRQV
 101 SAGNIGERL MATGARHYGD IRATAKWLE EVELPARIID DUPTFSGGM
 151 QORLOIANL VTHPKLVFMD EPTGLDVSV QARLLDLRG LVEELNAV
 201 IYTHDLGYAR LMDRLLYMK QGVVESGLT DVLDPBHPH YQLLVSVL
 251 ON

11AA_SEQUENCE 1.0 STANDARD: PRT: 263 AA.
 ID P1A_ORISA
 AC 040708;
 DT 01-NOV-1987 (Rel. 35, Created)
 DT 01-NOV-1987 (Rel. 35, Last sequence update)
 DT 01-NOV-1987 (Rel. 35, Last annotation update)
 DE PIR/A PROTEIN.
 DE Oryza sativa (Rice).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Oryza.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. INDICA-IR36;
 RA REIMANN C., MAUCH F., DUDLER R., HOFMANN C.;
 RT "Characterization of a rice gene induced by *Pseudomonas syringae* pv.
 RT syringae: requirement for the bacterial *lema* gene function.";
 RL Physiol. Mol. Plant Pathol. 46:71-81(1995).
 CC -1- SIMILARITY: STRONG, TO (S)-ACETONE-CYANOHYDRIN LYASE (EC
 CC 4.2.1.39).
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z34271; CA84025.1;
 DR HSBP; P52704; IYAS.
 DR PFAM: PF00561; abhydrolase_1.
 CC Hydrolyase; Serine esterase.
 FT ACT_SITE 82 82 BY SIMILARITY.
 FT ACT_SITE 213 213 BY SIMILARITY.
 FT ACT_SITE 241 241 BY SIMILARITY.
 SQ SEQUENCE 263 AA: 28663 MW; 0B95B213 CRC32;
 P1A_ORISA Length: 263 February 14, 2000 08:02 Type: P Check: 1305 ..

1 MEDGGHNEVF VHGLGYGAMC WYRVAAALRA AGHRAMALDM AAAGAHPARA
 51 DEVGSLSEYS RPLDDAVAAA AGERLVYVG HSLGLSLAL AMERFPDKVA
 101 AAVFLAACMP AAGKMGITL EEFMRIRKPD EFMDSKTIYL NTNQEPRTAV
 151 LGPKLAEK LYNSPPEDL TLATMLVRG TNYIDDPIMK DETLLEGVNY
 201 GSVKRYFLVA MDAASDEEM QRWITDLSRG VEVELAGAD HMAKCKPRE

251 LCDLLIRIAA KYD

11AA_SEQUENCE 1.0
ID PLAS_SYNY3 STANDARD; PRT; 126 AA.
AC P21697;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE PLASTOCYANIN PRECURSOR.
GN PETE OR SL0109.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91338701.
RA BRIGGS L.M., PECORARO V.L., MCINTOSH L.;
RT "Copper-induced expression, cloning, and regulatory studies of the
RT plastocyanin gene from the cyanobacterium Synechocystis sp. PCC
RT 6803.";
RL Plant Mol. Biol. 15:633-642(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE OF 29-47.
RX MEDLINE: 97443974.
RA SAZUKA T., OHARA O.;
RT "Towards a proteome project of cyanobacterium Synechocystis sp.
RT strain PCC6803: linking 130 protein spots with their respective
RT genes.";
RL Electrophoresis 18:1252-1258(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
RX MEDLINE: 98128796.
RA ROMERO A., DE LA CERDA B., VARELA P.F., NAVARRO J.A., HERVAS M.,
RA DE LA ROSA M.A.;
RT "The 2.15 A crystal structure of a triple mutant plastocyanin from
RT the cyanobacterium Synechocystis sp. PCC 6803.";
RL J. Mol. Biol. 275:327-336(1998).
CC -1- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN
CC -1- P700 AND THE CYTOCHROME B₆/F COMPLEX IN PHOTOSYSTEM I.
CC -1- INDUCTION: BY COPPER.
CC -1- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X54105; CAA38038.1; -
DR EMBL: D64000; BAA10227.1; -
DR PIR: S13733; S13733.
DR PDB: 1PCS; 17-DEC-97.
DR PROSITE: PS00196; COPPER-BIND. 1.
DR PFAM: PF00127; copper-bind. 1.
KW Electron transport; Copper; Signal; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 1 28
FT DOMAIN 29 126 PLASTOCYANIN-
FT METAL 67 67 PLASTOCYANIN-LIKE.
FT METAL 111 111 COPPER.
FT METAL 114 114 COPPER.
FT METAL 119 119 COPPER.
SQ SEQUENCE 126 AA; 13146 MW; 0358BF1 CRC32;

PLAS_SYNY3 Length: 126 February 14, 2000 08:02 Type: P Check: 6628 ..
1 MSKRLITLIL GLLVYSFF LSVSPAAN AYKMSDSG ALVEPSTVT
51 IKAGEYVWV NKLSPHNIV FAQGVADT AALSHKGLA PAGESTST
101 FPEPTIYVY CEPHAGAVY GKVWE
11AA_SEQUENCE 1.0
ID PLEX_HUMAN STANDARD; PRT; 350 AA.
AC P08567;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, last sequence update)
DT 15-DEC-1999 (Rel. 39, last annotation update)
DE PLECKSTRIN (PLATELET P47 PROTEIN).
GN PLEX OR P47.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88232910.
RA TYERS M., RACHUBINSKI R.A., MCCAW M.L., VARRICHIO A.M.,
RA SHOR R.G.L., HASLAM R.J., HARLEY C.B.;
RT "Molecular cloning and expression of the major protein kinase C
RT substrate of platelets.";
RL Nature 333:470-473(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89359547.
RA TYERS M., HASLAM R.J., RACHUBINSKI R.A., HARLEY C.B.;
RT "Molecular analysis of pleckstrin: the major protein kinase C
RT substrate of platelets.";
RL J. Cell. Biochem. 40:133-145(1989).
RN [3]
RP STRUCTURE BY NMR OF 1-105.
RX MEDLINE: 94268557.
RA YOON H.S., HADJUK P.J., PETROS A.M., OLEJNICZAK E.T., MEADOWS R.P.,
RA FESIK S.W.;
RT "Solution structure of a pleckstrin-homology domain.";
RL Nature 369:672-675(1994).
CC -1- FUNCTION: MAJOR PROTEIN KINASE C SUBSTRATE OF PLATELETS, ITS
CC EXACT FUNCTION IS NOT KNOWN.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07743; CAA30564.1; -
DR PIR: S00755; S00755.
DR PIR: A45762; A45762.
DR PDB: 1PSS; 03-JUN-95.
DR MTM: 173570; -
DR PROSITE: PS0003; PH DOMAIN; 2.
DR PFAM: PF00169; PH; 2.
DR PFAM: PF00610; DEP; 1.
KW Phosphorylation; Repeat; 3D-structure.
FT DOMAIN 107 120
FT CA_BIND 301 312 POTENTIAL.
FT DOMAIN 4 101 PH.
FT DOMAIN 244 347 PH.
FT VARIANT 92 92 W->R.
SQ SEQUENCE 350 AA; 40082 MW; 04F9FC1F CRC32;
PLEX_HUMAN Length: 350 February 14, 2000 08:02 Type: P Check: 1536 ..

1 MEKRRREGY LVKKSVENT WKPMVYLE DGEFFKKS DNSPKMPL
 51 KSLTSPQ DFKRMFEK ITTKQDHF FQALBERD AMVRLINKAI
 101 KCEGQKRA KSTRRSIRL PETIDGALY LSMTEKGI KEINLEKDK
 151 IFNHCTGNC VIDMLVSNOS VRRROGICMI ASSLINEGL QPAGDSKSA
 201 VDGTAENPFL DNPDAFYFP DSGFCEBNS SDDVILKEE FRYVILKGC
 251 LKQGHRRRN WKVKFIRE DPAYLYHPD AGAEDPLCAI HLRCVTVS
 301 ESNNSGRKE EENLEIITA DEHYFLQAA TPKEFTEMIK AIOMASFTGK

11AA_SEQUENCE 1.0
 ID PMA_MPCPO STANDARD: PRT: 299 AA.
 AC Q12639;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PECTIN LYASE PRECURSOR (EC 4.2.2.10).
 GN PEIA.

OS Mycosphaerella pinodes (Pea foot rot fungus).
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;
 OC Dothideales; Mycosphaerellaceae; Mycosphaerella.
 RN (1)
 RP SOURCE: FROM N.A.
 RC STRAIN-DSM 62763 / BERKELEY BLOXOM;
 RA HEIM P.;
 CC Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTIN TO GIVE
 CC OLIGOSACCHARIDES WITH TERMINAL 4-DEOXY-6-METHYL-ALPHA-D-GALACTO-
 CC 4-EPURONOSYL GROUPS.
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL)
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: X87580; CAA60884.1;
 DR PFAM: PF00544; pec_lyase; 1.
 KW Lyase; signal.
 FT SIGNAL 1 18
 FT CHAIN 19 299
 FT SIGNAL 19 299
 SO SEQUENCE 299 AA; 33072 MW; 121CF6C4 CRC32;

PLVA_MPCPO Length: 299 February 14, 2000 08:02 Type: P Check: 8743

1 MKSTFVSUG LTAITALAP TFSMDVKRD VYKRAASLSD VAIERPLRC
 51 RPSASLEPPS RYTRLEPSS LVLSRALATT SRAATSBS DRTARSLRT
 101 SPDSRVSRYM LSFVTLPSSR SVAMLSBSR RIELVTPKEI TTYLHRLQSE
 151 NWVWDLILS SDPDHDKDY DGLIDITHAA DEVTYNTFL HDHWKASLIG
 201 HSDSNGAEDK GHLVTVANN YLKNLSRGR PSASAGATST TTTTRCHMV
 251 STPARVSCL FTTFLSAPR RLFTRLMDM LSLATSVYV ARTLPRVL

11AA_SEQUENCE 1.0
 ID PMA_ASPAC STANDARD: PRT: 331 AA.
 AC Q12535;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PECTINESTERASE PRECURSOR (EC 3.1.1.11) (PECTIN METHYLESTERASE) (PE).
 GN PEI.

GN PEI.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-KSM 510;
 RX MEDLINE: 9079238.
 RA CHRISTGAT S., KOFOED L.V., HALKIER T., ANDERSEN L.N., HOCKAUF M.,
 RA DORREICH K., DAUBORG H., KAUPPINEN S.;
 RT "Pectin methyl esterase from Aspergillus aculeatus: expression,
 RT cloning in yeast and characterization of the recombinant enzyme."
 RL Biochem. J. 319:705-712(1996)
 CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: PECTIN N H(2)O -> N-METHANOL + PECTATE.
 CC -1- SIMILARITY: BELONGS TO THE PECTINESTERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: U49378; AAB42153.1;
 DR PROSITE: PS00800; PECTINESTERASE_1; 1.
 DR PROSITE: PS00503; PECTINESTERASE_2; 1.
 DR PFAM: PF01095; Pectinesterase; 1.
 KW Hydrolase; Serine esterase; Cell wall; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 331
 FT SIGNAL 17 331
 SO SEQUENCE 331 AA; 35681 MW; C567C2B9 CRC32;

PME_ASPAC Length: 331 February 14, 2000 08:02 Type: P Check: 2447

1 MKSVLAAL FVVSALAASR TPAISGATV AKSGDYTTI GAIDALSTS
 51 TTDITITIE EGYDEYVL PAMTGKITYI GQENTDSYA DNLVITTHAI
 101 SYEDAGESD LTRATFNKAV GSOVYNLINA NTGQACHQA LALSAMADQ
 151 GYGCNFTGY QDTLLAOTGN QLYINSYBEG AVDEFQCHA RAMFQNVDIR
 201 VVEGPTASAI TANGRSSETD TSYVYINKST VAAKSGDVA EGYLYGRFW
 251 SEYARVFOQ TSMTNVINSI GNTWSTSTP NTEVTFGEY ANTGAGEGT
 301 RASFAEKIDA KLITDILGS DYTSMVDISY F

11AA_SEQUENCE 1.0
 ID PME_ASPAC STANDARD: PRT: 331 AA.
 AC P17872;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PECTINESTERASE PRECURSOR (EC 3.1.1.11) (PECTIN METHYLESTERASE) (PE).
 GN PME1.

OS Aspergillus tubingensis.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-RH 5344;
 RX MEDLINE: 90332436.
 RA KHANH N.Q., ALBRECHT H., RUTKOWSKI E., LOEFFLER F., GOTTSCALK M.,
 RA JANY K.-D.;
 RT "Nucleotide and derived amino acid sequence of a pectinesterase cDNA
 RT isolated from Aspergillus niger strain RH 5344."
 RL Nucleic Acids Res. 18:4262-4262(1990).
 GN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92039066.

RA KHANH N.O., RUTKOWSKI E., LEIDINGER K., ALBRECHT H., GOTSCHALK M.;
 RT "Characterization and expression of a genomic pectin methyl esterase-
 RL gene 106:71-77(1991).
 CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: PECTIN + N H₂O = N METHANO + PECTATE.
 CC -1- SIMILARITY: BELONGS TO THE PECTINESTERASE FAMILY.
 CC -1- CAUTION: STRAIN RH 5344 WAS PREVIOUSLY SAID TO BE FROM A NIGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X52902; CAA37084.1; -
 DR EMBL: X54145; CAA38084.1; -
 DR PIR: S10487; S10487.
 DR JIR: J70589; J70589.
 DR PROSITE: PS00800; PECTINESTERASE_1; 1.
 DR PROSITE: PS00503; PECTINESTERASE_2; 1.
 DR PFAM: PF01095; Pectinesterase; 1.
 KW Hydrolyase; Serine esterase; Cell wall; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 331 PECTINESTERASE.
 SQ SEQUENCE 331 AA; 35715 MW; 80A4A5F0 CRC32;
 PHE_ASPTU Length: 331 February 14, 2000 08:02 Type: P Check: 609 ..

1 MKKSIASVL FAATALAASR MTPASGAIV AKSGDVTI SAADVLSLT
 51 STETQIFIE EGSYDEQYI PALSGLIYV GQEDITTYT SNLVNTHAI
 101 ALADVNDDE TATLRNAEG SAIYNINIAN TCGACRHAL AVASAYSEGG
 151 YVACFTGQ DTLAETGQ VYAGTIEGA VDFIFGQAR AMHEDIRV
 201 LGGPSASIT ANGRSESD SYIVHKSTV AADGNDVSS GYILGRPS
 251 QYARVCFQKT SMIVINHLG WTEWSTSTN TENVFEVYG NGTGAEGPR
 301 ANFSELTPE ITISWLGSD WEDWDTSYI N
 11AA_SEQUENCE 1.0
 ID PSPB_BOVIN STANDARD; PRT; 79 AA.
 AC P15781;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-NOV-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN B (SP-B) (6 KD PROTEIN)
 DE (PULMONARY SURFACTANT-ASSOCIATED PROTEOLIPID SP(L,PHE)).
 GN SPFB OR SPFB3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP MEDLINE: 88077030.
 RA OLAFSON R.W., RINK U., KIELAND S., YU S.-H., CHUNG J.,
 RA HARDING P.G.R., POSSWAYER F.;
 RT "Protein sequence analysis studies on the low molecular weight
 RT hydrophobic proteins associated with bovine pulmonary surfactant."
 RL Biochem. Biophys. Res. Commun. 148:1406-1411(1987).
 RN [2]
 RP MEDLINE: 88025156.
 RA YU S.-H., CHUNG W., OLAFSON R.W., HARDING P.G.R., POSSWAYER F.;
 RT "Characterization of the small hydrophobic proteins associated with
 RT pulmonary surfactant."
 RL Biochim. Biophys. Acta 921:437-448(1987).

CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
 CC AVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES
 CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINERTONS
 CC PER METER.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 DR PIR: A29667; A29667.
 DR PIR: S02317; S02317.
 KW Surface film; Gaseous exchange.
 SQ SEQUENCE 79 AA; 8660 MW; 2B73807E CRC32;
 PSPB_BOVIN Length: 79 February 14, 2000 08:02 Type: P Check: 7861 ..

1 FPIPIYCW LRLIKKIOA VIPKGYLMT VAQVCHVPL LVGSIIOOLV
 51 IEYSVILXTD TLGLRPNLV CGLRLRCSG
 11AA_SEQUENCE 1.0
 ID PTB_BACSU STANDARD; PRT; 299 AA.
 AC P54530;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROBABLE PHOSPHATE BUTYRYLTRANSFERASE (EC 2.3.1.19)
 DE (PHOSPHOTRANSBUTYRYLASE).
 GN Y01S
 OS Bacillus subtilis
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC KOBAYASHI Y., MIZONO M., MASUDA S., TAKEMARU K., HOSONO S.,
 RA SATO T., TAKEUCHI M.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CATALYSES THE CONVERSION OF BUTYRYL-COA THROUGH BUTYRYL
 CC PHOSPHATE TO BUTYRATE.
 CC -1- CATALYTIC ACTIVITY: BUTANOYL-COA + ORTHOPHOSPHATE = COA +
 CC BUTANOYLPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHATE ACETYLTRANSFERASE AND
 CC BUTYRYLTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D84432; BA12594.1; -
 DR EMBL: Z99116; CAB14340.1; -
 DR SUBTILIST; BG11722; Y01S.
 DR PFAM: PF01515; PTA_PTB; 1.
 KW Hypothetical protein; Transferase; Acyltransferase.
 SQ SEQUENCE 299 AA; 31772 MW; EF3F522D CRC32;
 PTB_BACSU Length: 299 February 14, 2000 08:02 Type: P Check: 158 ..

1 MKLKIDIGKA SIHKKTIAY AHAEDEVIR AVKLAHEILS APRLLTGDSK
 51 KLNELTSSMQ GHOVEIVHAN TPESAKLAV RAVHKRTADV LMGVNPVTSV
 101 LKAVINRQE GRSASYLSH VAVFDIPDP RLMEVTSAM NINPSLEELR
 151 QIQNNVHVA HAVGNMKA ALAAVEYVA PKMEIVYAA ALQMKKROO
 201 IKCITYDGL ALDNVSGIA AAKKISGVV AGNDILLVP TTAGNITLK

251 SLIFAKASY AAVITGAKAP IALTSRADSA ENKLYSIALA ICASEETH

11AA_SEQUENCE 1.0 STANDARD: PRT: 376 AA.
 ID PT16_HUMAN
 AC P35237;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 01-OCT-1996 (Rel. 34, last annotation update)
 DE PLACENTAL THROMBIN INHIBITOR (CYTOSOLASMIC ANTI-PROTEINASE) (CAP)
 DE (PROTEINASE INHIBITOR 5).
 GN P16 OR P11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 94022386.
 RA COUGHLIN P., SUN J., CERRUTI L., SALEM H.H., BIRD P.;
 RT "Cloning and molecular characterization of a human intracellular
 RT serine proteinase inhibitor.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:9417-9421(1993).
 RL [2]
 RN SEQUENCE FROM N.A. AND SEQUENCE OF 67-73 AND 144-149.
 RP TISSUE-PLACENTA;
 RX MEDLINE: 94183847.
 RA MORENSTERN K.A., SPRECHER C.A., HOLTH L., FOSTER D., GRANT F.J.,
 RA CHING A., KISIEL W.;
 RT "Complementary DNA cloning and kinetic characterization of a novel
 RT intracellular serine proteinase inhibitor: mechanism of action with
 RT trypsin and factor xa as model proteinases.";
 RT Biochemistry 33:3432-3441(1994).
 CC [1] SUBCELLULAR LOCATION: CYTOSOLASMIC.
 CC [1] TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, ALSO FOUND
 CC IN PLACENTA, CARDIAC MUSCLE, LUNG, LIVER, KIDNEY AND PANCREAS.
 CC [1] SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 222658; CA80373.1; -
 DR EMBL: 569272; AAB30320.1; -
 DR PIR: S35750; S35750.
 DR PIR: A48681; A48681.
 DR HSSP: P05619; 1HLE.
 DR MIM: 173321; -
 DR PROSITE: PS00284; SERPIN; 1.
 DR PFAM: PF00079; serpin; 1.
 KW Serpin. Serine protease inhibitor.
 FT ACT_SITE 341 342 REACTIVE BOND.
 FT ACT_SITE 175 175 G -> E (IN REF. 2).
 FT CONFLICT 362 362 R -> S (IN REF. 2).
 FT CONFLICT 376 376 R -> S (IN REF. 2).
 SO SEQUENCE 376 AA; 42587 MW; 2D806418 CRC32.

PT16_HUMAN Length: 376 February 14, 2000 08:02 Type: P Check: 300

1 MDLCEANGT FALNLKTLG KDNSKNVFFS PKMSICALAM VYMGAGNNA

51 AQAQALISLN KSGGGGDIHQ GFOSLLEVN KTGYOYLLRV ANLFGKSC

101 DFLSSRDSG OKFYQAEKEE LDFISAVEKS RKHINTVAE KTGKIAELL

151 SPGSVDPLTR LVLYNAVYFR GNDGQFDKE NTEERLEFVS KNEKPYQMA

201 FKQSTFKKTY IGEIFTQILV LPYVGKELNA IIMDPDEFTD LRTVEKELTY

251 EKFEVETRLD WDEDEVEEVS LPRFKLEST DMESVLRNLG MDAFELGKA

301 DFGMSQIDL SLKRVKSF VEVEEGTEA AATPAIMM RCARFVPRFC

11AA_SEQUENCE 1.0 STANDARD: PRT: 374 AA.
 ID PT18_HUMAN
 AC P50452;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE CYTOSOLASMIC ANTI-PROTEINASE 2 (CAP2) (PROTEINASE INHIBITOR 8).
 GN P18.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 96102039.
 RA SPRECHER C.A., MORENSTERN K.A., MATHEWS S., DAHLIN J.R.,
 RA SCHRADER S.K., FOSTER D.C., KISIEL W.;
 RT "Molecular cloning, expression, and partial characterization of two
 RT novel members of the ovalbumin family of serine proteinase
 RT inhibitors.";
 RT J. Biol. Chem. 270:29854-29861(1995).
 RL J. Biol. Chem. 270:29854-29861(1995).
 CC [1] SUBCELLULAR LOCATION: CYTOSOLASMIC.
 CC [1] SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: I40377; AAC41939.1; -
 DR HSSP: P05619; 1HLE.
 DR MIM: 601697; -
 DR PROSITE: PS00284; SERPIN; 1.
 DR PFAM: PF00079; serpin; 1.
 KW Serpin. Serine protease inhibitor.
 FT ACT_SITE 339 340 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 374 374 R -> S (IN REF. 2).
 SO SEQUENCE 374 AA; 42786 MW; 79AF74BC CRC32.

PT18_HUMAN Length: 374 February 14, 2000 08:02 Type: P Check: 4145

1 MDLCEANGT FALSLEKILG EDNSNRVFF SPMSISSALA WFMGAKGST

51 AQAQALISLN KYDGDHNGF QSLSEVNRT GTGYOYLLRVAN RLFGKTCDF

101 LPDFREYCKR FYQAELEELS FAEDTECRK HINDVVAEKT EKGISEVLDA

151 GTVDPRLKLV LVNAVYFKRG WNEQDFRKYT RGNLFTNEE KKYVMMEKE

201 AKFKMGYADE VHTQVLELPY VEELLSMVL LPDDNTDLAV VEKALYERK

251 KAMNSEKLT KSKVQYFLPR LKLESYDLE PELRLGMDI AFDEKADFS

301 GMSTERNVPL SKVAHKCFYE VNEEGTEAA AATVAVNSRC SRMEPRFCAD

351 HPFLFFIRRH KTNCLIFCR FSSP

11AA_SEQUENCE 1.0 STANDARD: PRT: 248 AA.
 ID PUR7_METH
 AC O26272;
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, last sequence update)
 DT 15-DEC-1999 (Rel. 39, last annotation update)
 DE PHOSPHORIBOSYLAMINOMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE (EC 6.3.2.6)
 DE (SAICAR SYNTHETASE).
 GN PURC OR MTH10.

OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H.
 RX MEDLINE: 98037514.
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DOBOIS J.,
 RA ALBREDE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTIER B., QIU D.,
 RA SPADAPORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PASHAKAR S.,
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.,
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delat: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7153(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + 1-(5-PHOSPHORIBOSYL)-4-CARBOXY-5-
 CC AMINOIMIDAZOLE + L-ASPARTATE = ADP + ORTHOPHOSPHATE + 1-(5-
 CC PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXYMIDATE)-5-AMINOIMIDAZOLE.
 CC -1- PATHWAY: SEVENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SAICAR SYNTHETASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-std.ch/announce/>
 CC or send an email to license@isb-std.ch).
 CC
 DR EMBL: AE000804; AAB84676.1;
 DR PROSITE: PS01057; SAICAR_SYNTHETASE_1; 1.
 DR PROSITE: PS01058; SAICAR_SYNTHETASE_2; 1.
 DR PFAM: PF01259; SAICAR_Syntc; 1.
 KW Purine biosynthesis; Ligase.
 SQ SEQUENCE 248 AA; 28244 MW; 05FA27E0 CRC32;
 PUR7_METH Length: 248 February 14, 2000 08:02 Type: P Check: 5550
 1 MDVKIDGPLY SGKAKDVLIT DDEIYAVRF RDTITADGE KKDTEMKCY
 51 YNSVISARIF EVLEAGVPT QYLELEPEGC ILARKLEMP IEVITENINAA
 101 GSIVRPEPFT EGQEFVPLI QMDYKSDENG DPMINDIIL ALGIATRDEL
 151 EIIRITLIHI NSVLRFILKS RGLILPDFKL EFGDSSGRI RLGDVSPPT
 201 CLRMEMGE PLDKIFRRG EESGVGAVNR VARILDED IEEMWNL
 11AA SEQUENCE 1.0 STANDARD; FRT; 310 AA.
 ID PYRB_ECOLI
 AC P00479; Q47555; Q47557
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1986 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE
 DE TRANS-CARBAMOYLASE) (ATCSE).
 GN PYRB.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84119419.
 RA SCHACHMAN H.K., PAUZA C.D., NAVRE M., KARELS M.J., WU L., YANG Y.R.;
 RT "Location of amino acid alterations in mutants of aspartate
 RT transcarbamoylase: structural aspects of interallelic
 RT complementation";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:115-119(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 83195078.
 RC STRAIN-K12

RA HOOVER T.A., ROOF W.D., FOLTERMANN K.F., O'DONOVAN G.A., BENICINI D.A.,
 RA WILD J.R.;
 RT "Nucleotide sequence of the structural gene (pyrB) that encodes the
 RT catalytic polypeptide of aspartate transcarbamoylase of Escherichia
 RT coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2462-2466(1983).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 95334362.
 RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
 RA BLATTNER F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [4]
 RP REVISION TO 195.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 9742617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RP SEQUENCE.
 RA KONIGSBERG W.H., HENDERSON L.;
 RT "Amino acid sequence of the catalytic subunit of aspartate
 RT transcarbamoylase from Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2467-2471(1983).
 RN [6]
 RP SEQUENCE OF 1-53 FROM N.A.
 RX MEDLINE: 89017155.
 RA ROLAND K.L., LIU C., TURNBOUGH C.L. JR.;
 RT Role of the ribosome in suppressing transcriptional termination at
 RT the pyri attenuator of Escherichia coli K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7149-7153(1988).
 RN [7]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 85289046.
 RA ROLAND K.L., POWELL F.E., TURNBOUGH C.L. JR.;
 RT "Role of translation and attenuation in the control of pyrBI operon
 RT expression in Escherichia coli K-12.";
 RL J. Bacteriol. 163:991-999(1985).
 RN [8]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 83169660.
 RA TURNBOUGH C.L. JR., HICKS K.L., DONAHUE J.P.;
 RT "Attenuation control of pyrBI operon expression in Escherichia coli
 RT K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:368-372(1983).
 RN [9]
 RP SEQUENCE OF 1-17 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 91035438.
 RA DONAHUE J.P., TURNBOUGH C.L. JR.;
 RT Characterization of transcriptional initiation from promoters p1 and
 RT p2 of the pyrBI operon of Escherichia coli K12.";
 RL J. Biol. Chem. 265:19091-19099(1990).
 RN [10]
 RP SEQUENCE OF 1-11.
 RC STRAIN-K12 / W3110;
 RA PASQUALE C., SANCHEZ J.-C., RAVIER F., GOLAZ O., HUGHES G.J.,
 RA FEUTIGER S., PAQUET N., WILKINS M., APPEL R.D., BAIRDOCH A.,
 RA HOCHSTRASSER D.F.;
 RT Submitted (Sep-1994) to the SWISS-PROT data bank.
 RN [11]
 RP SEQUENCE OF 1-20.
 RC STRAIN-K12 / EMG2;

RX MEDLINE: 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RT Electrophoresis 18:1259-1313(1997).
RN [12]
RP SEQUENCE OF 1-5.
RC STRAIN-K12 / W3110;
RX MEDLINE: 98291876;
RA MOLLOI M.P., HERBERT B.R., WALSH B.J., TYLER M.I., TRAINI M.,
RA SANCHEZ J.-C., HOCHSTRASSER D.F., WILLIAMS K.L., GOOTER A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RT Electrophoresis 19:837-844(1998).
RL [13]
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND SUBUNITS.
RP MEDLINE: 84248054.
RX KE H.-M., HONZARO R.B., LIPSCOMB W.N.;
RA "Structure of unligated aspartate carbamoyltransferase of Escherichia
RT coli at 2.6-A resolution.";
RT Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE: 91104727.
RA STEVENS R.C., GOUAUX J.E., LIPSCOMB W.N.;
RT "Structural consequences of effector binding to the T state of
RT aspartate carbamoyltransferase: crystal structures of the unligated
RT and ATP- and CTP-complexed enzymes at 2.6-A resolution.";
RT Biochemistry 29:7691-7701(1990).
RN [15]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE: 91104728.
RA GOUAUX J.E., STEVENS R.C., LIPSCOMB W.N.;
RT "Crystal structures of aspartate carbamoyltransferase ligated with
RT phosphonacetamide, malonate, and CTP or ATP at 2.8-A resolution and
RT neutral pH.";
RT Biochemistry 29:7702-7715(1990).
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYR5 CHAINS
CC ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PRTI CHAINS
CC ORGANIZED AS THREE DIMERS (R2).
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01472; AAA24476.1; -
DR EMBL: M10743; AAA24479.1; -
DR EMBL: J01670; AAA24474.1; -
DR EMBL: U14003; AAA97142.1; -
DR EMBL: AF000495; AAC77202.1; -
DR EMBL: M60508; AAA24461.1; -
DR PIR: A00561; DPCC
DR PIR: AACM; 15-JUL-92.
DR PDB: 2ATC; 15-JUL-92.
DR PDB: 8ATC; 15-OCT-90.
DR PDB: 1AT1; 15-OCT-90.
DR PDB: 2AT1; 15-OCT-90.
DR PDB: 3AT1; 15-OCT-90.
DR PDB: 4AT1; 15-OCT-90.
DR PDB: 5AT1; 31-JUL-94.
DR PDB: 6AT1; 15-OCT-90.
DR PDB: 7AT1; 15-OCT-90.
DR PDB: 8AT1; 31-JUL-94.
DR PDB: 1RAA; 31-JAN-94.
DR PDB: 1RAB; 31-JAN-94.

DR PDB: 1RAC; 31-JAN-94.
DR PDB: 1RAD; 31-JAN-94.
DR PDB: 1RAE; 31-JAN-94.
DR PDB: 1RAF; 31-JAN-94.
DR PDB: 1RAG; 31-JAN-94.
DR PDB: 1RAI; 31-JAN-94.
DR PDB: 1RAJ; 31-JAN-94.
DR PDB: 1RAK; 31-JAN-94.
DR PDB: 1RAB; 14-OCT-98.
DR SWISS-2DPAGE: P00479; COLI.
DR ECO2DBASE: H031.3; 6TH EDITION.
DR PROSITE: PS00997; CARBAMOYLTRANSFERASE; 1.
DR PFM: PFM0185; OTCASE; 1.
RW Pyrimidine biosynthesis; Transferase; 3D-structure.
FT INT_MET 0
FT CONFLICT 60
FT CONFLICT 86
FT CONFLICT 90
FT CONFLICT 129
FT CONFLICT 149
FT CONFLICT 195
FT CONFLICT 220
FT CONFLICT 256
FT CONFLICT 259
FT CONFLICT 296
FT TURN 3
FT STRAND 9
FT HELIX 12
FT HELIX 17
FT TURN 36
FT STRAND 43
FT HELIX 53
FT TURN 65
FT STRAND 69
FT TURN 81
FT HELIX 88
FT TURN 96
FT STRAND 101
FT TURN 109
FT HELIX 110
FT HELIX 111
FT TURN 117
FT TURN 120
FT STRAND 124
FT TURN 130
FT HELIX 135
FT STRAND 156
FT TURN 163
FT HELIX 167
FT TURN 177
FT STRAND 183
FT HELIX 190
FT TURN 196
FT STRAND 205
FT TURN 208
FT STRAND 212
FT TURN 215
FT HELIX 218
FT TURN 219
FT STRAND 224
FT HELIX 232
FT TURN 237
FT HELIX 240
FT TURN 241
FT STRAND 244
FT TURN 249
FT HELIX 251
FT TURN 255
FT TURN 259
FT STRAND 262
FT TURN 264
FT STRAND 273
FT HELIX 275
FT TURN 278
FT TURN 281
FT STRAND 283

Mon Feb 14 08:07:43 2000

sp.cat

Page 102

FT HELIX 285 304
SO SEQUENCE 310 AA; 34296 MW; BDF9E927 CRC32;
PYRB_ECOLI Length: 310 February 14, 2000 08:02 Type: P Check: 3945

1 ANPLYOKHII SINDLSRDL NLVLATAKL KANPOPELKH HKVIASCFEE
51 ASTRIRLSEF TSMHRLGASV VGFSDSANTS LGKKGTELD TIVISTYVD
101 AIVARHPOEG AARLATEFSG NVPVLNAGDG SNQHPQTLL DLFIOETOG
151 RLDELHVAHV GDLKRGRTV SLFOALAKFD GNFETIADY ALAPQYILD
201 MLDEKGIAMS LHSIEEVVA EVDILYMTV OKERLDPSEY ANVAQOYLR
251 ASDLHNKAKN MKVHLPLPRV DEITVDYKT PHAWTFQOAG NGIFARQALL
301 ALVLRDLYL

11AA_SEQUENCE 1.0
ID PYRB_MYCTU STANDARD; PRT; 319 AA.
AC P71808;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE
DE TRANSCARBAMYLASE) (ATCASE).
GN PYRB OR RV1380 OR MTCY02B12.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E., III, TERKLA F.,
RA BADOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORSLEY T., JAGELS K., KROGH A., MCLEAN J., MOUPEY S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RALANDHEM M.A., ROGERS J.,
RA RUTER S., SEGER K., SKELTON S., SQUARES S., SOKES R., SOUTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RT -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE -
CC -1- ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z81011; CAB02641.1; -
DR HSSP: P00479; 3AT1.
DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
KW PFAM: PF00185; OTCase; 1.
KW Pyrimidine biosynthesis; Transferrase.
SO SEQUENCE 319 AA; 33818 MW; 38CF9A86 CRC32;

PYRB_MYCTU Length: 319 February 14, 2000 08:02 Type: P Check: 1547

1 MTPHLLTAA DLSRDATAI LDDADRAQA LVGRDITKLP TLKGRIVYTM
51 FYENSTRTV SEVAGKMS ADVINVSAG SSVGKESIR DTAITLRAAG
101 ADALITHPA SGAAHLAOM TGAHNDGPV INAGDGTHER PTOALLDALI

151 IRQRLGIEG RIYIVGDL HSRVARSNM LDTLGAEVY LVAPETLLPV
201 GYTGMPATVS HPEDELPAQA DAVILMRVA ERMNGGFPS VREYSVRYGL
251 TERQAMLPQ HAVVILPGPM VRGMEITISV ADSOSAVILO QVSNQOVVRM
301 ALVFLVAVGA ODAKRGKAA

11AA_SEQUENCE 1.0
ID PYRB_PYRB STANDARD; PRT; 308 AA.
AC P77918;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE
DE TRANSCARBAMYLASE) (ATCASE).
GN PYRB.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GE5.
RX MEDLINE: 97352668.
RA PURCAREA C., HERVE G., LADJANI M.M., CUNIN R.,
RT "Aspartate transcarbamylase from the deep-sea hyperthermophilic
RT archaeon Pyrococcus abyssi: genetic organization, structure, and
RT expression in *Escherichia coli*."
RT J. Bacteriol. 179:4143-4157(1997).
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE -
CC -1- ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PRIMAR: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONTAINS CATALYTIC AND REGULATORY CHAINS.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U61765; AAB62984.1; -
DR HSSP: P00479; 3AT1.
DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
KW PFAM: PF00185; OTCase; 1.
KW Pyrimidine biosynthesis; Transferrase.
SO SEQUENCE 308 AA; 34901 MW; 7CEC7D77 CRC32;

PYRB_PYRB Length: 308 February 14, 2000 08:02 Type: P Check: 9304

1 MMKRDVVIS IRPSEKIDIE TVLATAELE REIKERGOLE YAKKILATL
51 FPEBTRRL SFSANHRIG GAVIGFAEAS TSSVKKGESL RDTIKTYEY
101 CDVYIIRHP EGARLADEV AVEPVINAGD GSNQHPQTLL LDLYTIKKEF
151 GRIDELKIGL LGDLKYGRTV HSLAEALTFY DVELYLISPE LLRMPRIIVE
201 ELREKMKVY ETTLEDVIG KLDVLYTRI OKERPEDEO YLKVGSGYOV
251 NLKYLAKAD ELRIMHPLR VDEIHPEVDN TKHATYFROV FNGVPVMAI
301 LALVIGVI

11AA_SEQUENCE 1.0
ID PYRB_PYRB STANDARD; PRT; 308 AA.
AC O58451;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE

DE TRANSCARBAMYLASE) (ATCASE).
 GN PYRB OR PH0720.
 OS Pyrococcus horikoshii.
 OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3.
 RX MEDLINE: 96344137.
 RA KAMARABAYATI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAKA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,
 RA FUNABASHI T., TANAKA T., KODOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
 RA AOKI K., NAKAMURA Y., ROBE T.F., HORIKOSHI K., MASUCHI Y., SHIZUHA H.,
 RA KIKUCHI H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
 CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
 CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
 CC -1- SUBUNIT: CONTAINS CATALYTIC AND REGULATORY CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AP000003; BAA28611.1;
 DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
 DR PFAM: PF00185; OTCase; 1; Transferase.
 KW Pyrimidine biosynthesis; 1263350D CRC32;
 SQ SEQUENCE 308 AA; 34862 MW; 1263350D CRC32;
 PYRB_PYRHO Length: 308 February 14, 2000 08:02 Type: P Check: 9980 ..
 1 MEMKGRDVIS INDPSKEDIE VVLSAERLE KEMKRGOLE YANKKILATI
 51 FEPSTRTRL SPESAMHRLG GSVIGFVAS TSVKKGESL RDIKTEVOY
 101 SDYIVIRHPK EGAARLAEV ADIVINAGD GSNQHPOTLL IDLYTIKKEF
 151 GTIDGLKIGL LGDLKGRIV HSLAEALAFY DVELYLISPE LLRMPKHIVE
 201 ELERGRMKIV ETTKLEEVIG ELVDLYYTRI QKERPEDEGE YLTKGSGYOV
 251 NKILEVAVD SLRIMHPLPR VDEIHPEVDK TKHAIFKQV FNGVPVBMAL
 301 LALVLYGI
 11AA_SEQUENCE 1.0 STANDARD; PRT; 310 AA.
 ID PYRB_SALTU
 AC P08420;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE
 DE TRANSCARBAMYLASE) (ATCASE).
 GN PYRB.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 87246692.
 RA MICHAELS G., KELIN R.A., NANGANG F.E.;
 RT "Cloning, nucleotide sequence and expression of the pyrBI operon of
 RT Salmonella typhimurium LT2.";
 RL Eur. J. Biochem. 166:55-61(1987).

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
 CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
 CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
 CC -1- SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS
 CC ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS
 CC ORGANIZED AS THREE DIMERS (R2).
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X03641; CAZ9129.1;
 DR PIR: S00479; OMEBAC.
 DR HSSP: P00479; IRAG.
 DR STYGENE; SG10324; PYRB.
 DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
 DR PFAM: PF00185; OTCase; 1; Transferase.
 KW Pyrimidine biosynthesis; 1.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 310 AA; 34307 MW; 01405B57 CRC32;
 PYRB_SALTU Length: 310 February 14, 2000 08:02 Type: P Check: 2583 ..
 1 ANPLVOKHII SINDLSRDL NLYLATAKL KANQPELLK HVIASCFEE
 51 ASTRRLSFE TSMHRLGASV VGFSDSANTS LGKRGETLAD TISVISTYVD
 101 AIVMHPQEG AARLATEFSG QVPVLNAGDG SNGHPOTLL DFTIOETOG
 151 RLDNHIIMV GDLKGRIVH FAKPRTIAKF SGNFYIAP DLAMPQYLL
 201 DMLDEKQAV SLHSISEVM ADVLLIYMR VQKRELPSE YANVAQFVL
 251 RPDNGAREN MKVHLPLRI DETTDVDTK PHANYFOAG NGIFAAQLL
 301 ALVNLSEL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 305 AA.
 ID PYRB_SERMA
 AC P19910;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE
 DE TRANSCARBAMYLASE) (ATCASE).
 GN Serratia marcescens.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Serratia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89380286.
 RA BECK D., KEDZIE K.M., WILD J.R.;
 RT "Comparison of the aspartate transcarbamoylases from Serratia
 RT marcescens and Escherichia coli.";
 RL J. Biol. Chem. 264:16629-16637(1989).
 CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
 CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
 CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
 CC -1- SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS
 CC ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS
 CC ORGANIZED AS THREE DIMERS (R2).
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: J05033; AAA26564.1; -

DR PIR: B34396; OMSEAC.

DR HSSP: P00479; 1RAG.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

DR PFAM: PF00185; OTCace; 1.

KW Pyrimidine biosynthesis; Transferase.

FT INIT MET 0 BY SIMILARITY

SEQUENCE 305 AA; 33240 MW; 3E4D883A CRC32;

PYRB_SERNA Length: 305 February 14, 2000 08:02 Type: P Check: 4826 ..

1 ANPLVHKHII SINDISRDL ELVLAPAGI KANPOPELLK HVIYASCFEE

51 ASTRTLSFE TSMHRLGASV VGFADGNTS LKKKGETLAD TISVISTYVD

101 AIYMRHPQEG ARMASESGN VPVLNAGDN OHPITLIDL FTIOETQGR

151 SNLSIAMVGD LKGRVYHSL TOALAFEGN RFFYAPDAL AMPAYILKML

201 EEKGIYESSH GSIEVYVPEL DIYYMTRVOK ERLDPSEYAN VKAQVYLAAD

251 LAGANMLKVL HPLPRIDEIA TDVDTKPHAY YFOQAGNGIF ARSALALVYN

301 ADLAL

11AA_SEQUENCE 1.0

ID PYRB_VIBS2 STANDARD; PRT; 310 AA.

AC P96174;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 38, Last annotation update)

DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE TRANSFERASE) (ATCASE).

GN PYRB.

OS Vibrio sp. (strain 2693).

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DE CASTELE M., LIANG Z., FENG Z.Y., LEGRAND C., GLANDORFF N.; Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE - ORTHOPHOSPHATE + N-CARBAMOYL-ASPARTATE.

CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SUBUNIT: CONTAINS SIX CATALYTIC AND SIX REGULATORY CHAINS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Y09786; CAAT0923.1; -

DR HSSP: P00479; 1RAG.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

DR PFAM: PF00185; OTCace; 1.

KW Pyrimidine biosynthesis; Transferase.

SO SEQUENCE 310 AA; 34419 MW; C460D24E CRC32;

PYRB_VIBS2 Length: 310 February 14, 2000 08:02 Type: P Check: 5096 ..

1 MANPLFRKHI VSINDISRNE LELIVYTAAK LKKOPEPELL KKNVIASCFE

51 EASTRTLSFE ETAIORLGT VIGFDNASVT SLAKKGETLA DSISVSYV

101 DAFVRRHPOE GAARLASEFS NVPVINGDGD SNQPTPTIL DFISETQGC

151 CLDNLNIALV GDLKGRVTH SLAQALAKFS GCKFYFIAPD ALAMPEYICD

201 ELDEHNYSYA CYSIEEYVP EIDVLYMTRV QKRFPEFETX QHKAAGFLTS

251 ASSLKHAQDN LKYLHPLPRV DEIVADVDT PYAYYFOAE NGVAREALL

301 ALVYNATIEG

11AA_SEQUENCE 1.0

ID PYR2_BACSTU STANDARD; PRT; 256 AA.

AC P25983;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE DIHYDROOrotate DEHYDROGENASE ELECTRON TRANSFER SUBUNIT.

GN PYRDI OR PYRDB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RA OUTIN C.L., STEPHENSON B.T., SWITZER R.L.; "Functional organization and nucleotide sequence of the Bacillus subtilis pyrimidine biosynthetic operon.";

RT J. Biol. Chem. 266:9113-9127(1991).

RN [2]

RP FUNCTION.

RX MEDLINE: 96326349.

RA KÄHLER A.E. SWITZER R.L.; "Identification of a novel gene of pyrimidine nucleotide biosynthesis, pyrdii, that is required for dihydroorotate dehydrogenase activity in Bacillus subtilis.";

RL J. Bacteriol. 178:5013-5016(1996).

CC -1- FUNCTION: PROBABLE ELECTRON CARRIER PROTEIN INVOLVED IN THE TRANSFER OF REDUCING EQUIVALENTS FROM THE FLAVOPROTEIN SUBUNIT TO THE ELECTRON TRANSPORT SYSTEM IN THE CELL MEMBRANE.

CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SIMILARITY: TO S. TYPHIMORIUM ASRB AND P. FURIOSUS HYDG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M59757; AAA21271.1; -

DR EMBL: Z09112; CAB13427.1; -

DR PIR: G39845; G39845.

DR SUBTILISY: HG10727; PYRDI.

KW Pyrimidine biosynthesis; Electron transport.

SO SEQUENCE 256 AA; 28099 MW; 3F84D3B CRC32;

PYR2_BACSTU Length: 256 February 14, 2000 08:02 Type: P Check: 2658 ..

1 MKRAYLTVCN NOQIADRVFQ MWLKGELVGC FTTPOQFLHL KYSEAVTPLL

51 RRPISADYVN FEKNEVITII RVDEGTRLL SLKQOGEIVD VLSPUNGFPF

101 VNEVOPGKTA LLYGGVGVP PLOELSKRLI EKGAVVIVHL GFOSAKDVY

151 EEECRQYGDY YVATADGSYG ETGEVIVDVIK RKLEFDILL SCGTPMLKA

201 LKQYAHKEV YLSMERMG GIGACFACVC HTNESETSYV KYCIDGVYFK

251 AQEVAL

11AA_SEQUENCE 1.0

ID RAS1_DROME STANDARD; PRT; 336 AA.

AC Q27257;

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-DEC-1999 (Rel. 39, Last annotation update)
 DE DNA REPAIR PROTEIN RAD51 HOMOLOG (RECA PROTEIN HOMOLOG).
 GN RAD51 OR DMR.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S:
 RX MEDLINE: 95161094.
 RA AKABOSHI E., INOUE Y., RYO H.:
 RT "Cloning of the cDNA and genomic DNA that correspond to the reca-like
 RT gene of Drosophila melanogaster."
 RL Jpn. J. Genet. 69:663-670(1994).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R:
 RX MEDLINE: 96207535.
 RA MCKEE B.D., REN X.J., HONG C.S.:
 RT "A reca-like gene in Drosophila melanogaster that is expressed at
 RT high levels in female but not male meiotic tissues."
 RL Chromosoma 104:479-488(1996).
 CC -1- FUNCTION: BINDS TO SINGLE AND DOUBLE STRANDED DNA AND EXHIBITS
 CC DNA-DEPENDENT ATPASE ACTIVITY. UNDERWINDS DUPLEX DNA (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 CC PROKARYOTIC RECA PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL: D37788; BAA07039.1; -
 DR EMBL: D17725; BAA04580.1; -
 DR EMBL: L41342; AAA64873.1; -
 DR FJBASE: FB90011700; Rad51.
 KW DNA-binding; ATP-binding; Nuclear protein.
 FT NP-BIND 124 131
 FT NP-BIND 124 131
 SQ SEQUENCE 336 AA; 36647 MW; 5BDE65D2 CRC32;
 RA51_DROME Length: 336 February 14, 2000 08:02 Type: P Check: 2333 ..
 1 MEKLTINVOAQ QEEEEEGL SVTKLIGSI TARDIKLQO ASLHTVESVA
 51 NATKQIMAI PGIAGGKVEO IITEANKLVP LGFLSARTEY QMRADYVOLS
 101 TGSKELDKLL GGGIETGSIY EIFGEFRCKG TOLCHTLAVY COLPIQKRG
 151 EGRKCMYIDTE NTFRPERLAA IAQRKLNES EVDNVAFTF AHNSDQTKL
 201 IOMAGMLFE SRVALIVDS AMALYRSDYI GRGELARON HGLFLRMLO
 251 RLADFEQAV VITNQTASL DGAPGMFAK KPIGHIHMA SSTRLYLAK
 301 GKGETRICKI YDSPCLPESB AMFALLPGCI GDARES
 11AA_SEQUENCE 1.0
 ID RBRC CHRVI STANDARD; PRT; 302 AA.
 AC P25544;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE RUBISCO OPERON TRANSCRIPTIONAL REGULATOR.
 DE RUBISCO
 GN Chromatium vinosum.

OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
 OC Allochromatium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91317745.
 RA VIALE A.M., KOBAYASHI H., AKAZAWA T., HENIKOFF S.:
 RT "rbcR, a gene coding for a member of the lysR family of
 RT transcriptional regulators, is located upstream of the expressed set
 RT of ribulose 1,5-bisphosphate carboxylase/oxygenase genes in the
 RT photosynthetic bacterium Chromatium vinosum."
 RL J. Bacteriol. 173:5224-5229(1991).
 RN [2]
 RP SEQUENCE OF 1-57 FROM N.A.
 RX MEDLINE: 89213919.
 RA VIALE A.M., KOBAYASHI H., AKAZAWA T.:
 RT "Expressed genes for plant-type ribulose 1,5-bisphosphate
 RT carboxylase/oxygenase in the photosynthetic bacterium Chromatium
 RT vinosum" which possesses two complete sets of the genes."
 RL J. Bacteriol. 171:2391-2400(1989).
 CC -1- FUNCTION: TRANS-ACTING TRANSCRIPTIONAL REGULATOR OF RUBISCO
 CC -1- GENES (RBCAB) EXPRESSION.
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL: M64032; AAA23327.1; -
 DR EMBL: M26396; -; NOT_ANNOTATED_CDS.
 DR PIR: A40369; A40369.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY.1.
 DR PIRAM: PF00126; HTH.1.
 KW Transcription regulation; Activator; DNA-binding.
 FT DNA-BIND 20 39
 FT H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 302 AA; 34092 MW; CA38CF20 CRC32;
 RBCK CHRVI Length: 302 February 14, 2000 08:02 Type: P Check: 4295 ..
 1 MHVSLQKAV FEAVARHNSY TRAAEHLHS QPAVEMOYRQ LEDDIGSLF
 51 ERGKQVLT ENGREVFHS RAIGQSIREM EVELSLKGV SRGSLRAVA
 101 STVNFAPRL MAIFQHRSG IGLRDLVTR ESYVMDLSN SVDLYMGVP
 151 PRNVEVEAER FMDNPLVIA PPDHPLAGER AISLARLAE TFWRECGSG
 201 TROMERFES ERGQIRHGM QMTRNEAVKO AVRSGLGLSV VSLHTLEL
 251 ETRRLVTLDV EGFPDRQWY LVYRGRKRLS PAAGAFRETV LSEARMHCR
 301 LG
 11AA_SEQUENCE 1.0
 ID RBSO SOLTU STANDARD; PRT; 181 AA.
 AC P10647;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN C PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT C).
 GN RBS-C.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. HH1201/7;
 RX MEDLINE: 88124937;
 RA WOLTER F.P., FRITZ C.C., WILMITZER L., SCHELL J., SCHREIER P.H.;
 RT "rbcs genes in Solanum tuberosum: conservation of transit peptide and
 exon shuffling during evolution.";
 RL Proc Natl Acad Sci U S A. 85:846-850(1988).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 D-RIBULOSE 1,5-BISPHOSPHATE. THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 2 3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: J03613; AAA33838.1; -
 DR PIR: A31083; RPOSC.
 DR HSSP: P00866; 1RLC.
 DR MENDEL: 15074; SOLTU:rbcs:tm15074.
 DR PRFM: PF00101; RUBISCO_small1.1
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 KW Multigene family; 58 CHLOROPLAST (BY SIMILARITY).
 FT TRANSLIT 59 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 FT CHAIN
 SO SEQUENCE 181 AA; 20368 MW; 44574D48 CRC32;
 RBSO_SOLTU Length: 181 February 14, 2000 08:02 Type: P Check: 4699 ..
 1 MASSIVSAA VATRSNVAQA SWAPFTGLK SAASEPYTKK NNNVDITSIA
 51 SNGGRVRCMQ VMPPIKMKY ETLSTPLDLS DEQLKEVEY LKNGWVPC
 101 EEFTEHGVY REHNSPGY DGRYWTMVKL PMFGCTDQTO VLAEOEAKN
 151 AYPQAWRII GFDNVRQVOC ISFIAYKPEG Y
 11AA_SEQUENCE 1.0
 ID RBS1_LYCSES STANDARD; PRT; 181 AA.
 AC P08706;
 DR 01-JAN-1988 (Rel. 06, Created)
 DT 01-NOV-1980 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 1) (LESS17).
 GN RBGS-1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; easterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A. (LESS 17).
 RC STRAIN-CV. VF36;
 RX MEDLINE: 87163513.
 RA MCKNIGHT T.D., ALEXANDER D.C., BABCOCK M.S., SIMPSON R.B.;
 RT "Nucleotide sequence and molecular evolution of two tomato genes
 encoding the small subunit of ribulose-1,5-bisphosphate
 carboxylase.";

RL Gene 48:23-32(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (RBGS-1).
 RX MEDLINE: 86233336.
 RA PICHERSKY E., BERNATZKY R., TANKSLEY S.D., CASHMORE A.R.;
 RT "Evidence for selection for a mechanism in the concerted evolution of
 Lycopersicon esculentum (tomato) genes encoding the small subunit of
 ribulose-1,5-bisphosphate carboxylase/oxygenase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3880-3884(1986).
 RN [3]
 RP SEQUENCE FROM N.A. (RBGS-1).
 RC STRAIN-CV. VENT CHERRY LA1221;
 RX MEDLINE: 86038372.
 RA SUGITA M., MANZARA T., PICHERSKY E., CASHMORE A., GRUISEM W.;
 RT "Genomic organization, sequence analysis and expression of all five
 genes encoding the small subunit of ribulose-1,5-bisphosphate
 carboxylase/oxygenase from tomato.";
 RL Mol. Gen. Genet. 209:247-256(1987).
 RN [4]
 RP SEQUENCE OF 1-20 FROM N.A. (RBGS-1).
 RC STRAIN-CV. VENT CHERRY LA1221; TISSUE=ROOT;
 RX MEDLINE: 93144693.
 RA MANZARA T., CARRASCO P., GRUISEM W.;
 RT "Developmental and organ-specific changes in DNA-protein interactions
 in the tomato rbcs1, rbcs2 and rbcs3a promoter regions.";
 RL Plant Mol. Biol. 21:69-88(1993).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 2 3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: M15235; AAA34191.1; -
 DR EMBL: M15542; AAA34188.1; -
 DR EMBL: X05982; CAA29400.1; -
 DR EMBL: X65068; CAA46868.1; -
 DR PIR: S02364; RKTOSI.
 DR HSSP: P00866; 1RLC.
 DR PRFM: PF00101; RUBISCO_small1.1
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 KW Multigene family; 58 CHLOROPLAST (BY SIMILARITY).
 FT TRANSLIT 59 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 FT CHAIN
 SO SEQUENCE 181 AA; 20307 MW; 981E9B3C CRC32;
 RBS1_LYCSES Length: 181 February 14, 2000 08:02 Type: P Check: 5240 ..
 1 MASSIVSAA AATRSNVAQA SWAPFTGLK SAASEPYTKK NNNVDITSIA
 51 SNGGRVRCMQ VMPPIKMKY ETLSTPLDLS DEQLSEIEY LKNGWVPC
 101 EEFTEHGVY REHNSPGY DGRYWTMVKL PMFGCTDQTO VLAEOEAKK
 151 AYPQAWRII GFDNVRQVOC ISFIAYKPEG F

11AA_SEQUENCE 1.0 STANDARD; PRT: 180 AA.
 ID RB51_PETSP
 AC P04714.1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN SS08 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT SS08).
 GN RB51.
 OS Petunia sp. (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Petunia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. MITCHELL;
 RX MEDLINE; 86205237.
 RA TURNER N.E., CLARK W.G., TABOR G.J., HIRONAKA C.M., FRALEY R.T.,
 SHAH D.M.;
 RT "The genes encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase are expressed differentially in petunia leaves."
 RT Nucleic Acids Res. 14:3325-3342(1986).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X03820; CAA27444.1; -
 DR PIR: A24917; RKPS8.
 DR HSSP: P00866; 1RJC.
 DR MENDEL: 15078; PETSP; rbcS; mnl5078.
 DR PFM: PFO0101; RUBISCO_small; 1.
 DR PFM: PFO0101; RUBISCO_small; 1.
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 KM Multigene family.
 FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 58 180 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 FT SO SEQUENCE 180 AA; 20370 MW; 71E5C2F3 CRC32;
 RB51_PETSP Length: 180 February 14, 2000 08:02 Type: P Check: 2967 ..

01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 1).
 GN RB51.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HH1201/7;
 RX MEDLINE; 8812437.
 RA WOLTER F.P., FRITZ C.C., WILMITZER L., SCHELL J., SCHREIER P.H.;
 RT "rbcS genes in Solanum tuberosum: conservation of transit peptide and
 RT exon shuffling during evolution."
 RT Proc. Natl. Acad. Sci. U.S.A. 85:846-850(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. AM 80.5793;
 RA FRITZ C.C., WOLTER F.P., SCHENKMEYER V., HERGET T., SCHREIER P.H.;
 RT submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X69759; CAA49413.1; -
 DR PIR: B31083; RKPOS1.
 DR PIR: S31497; S31497.
 DR HSSP: P00866; 1RJC.
 DR MENDEL: 490; SOLTU; rbcS; 1.
 DR PFM: PFO0101; RUBISCO_small; 1.
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 KM Multigene family.
 FT TRANSIT 1 58 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 59 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 FT SO SEQUENCE 181 AA; 20556 MW; 5EA0260B CRC32;
 RB51_SOLTU Length: 181 February 14, 2000 08:02 Type: P Check: 4863 ..

11AA_SEQUENCE 1.0 STANDARD; PRT: 181 AA.
 ID RB52_SOLTU
 AC P26574.1
 DT 01-AUG-1992 (Rel. 23, Created)

11AA_SEQUENCE 1.0 STANDARD; PRT: 180 AA.
 ID RB52_LYCES
 AC P07179.1
 DT 01-APR-1988 (Rel. 07, Created)

DT 01-NOV-1990 (Rel. 16, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE RUBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN 2A PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 2A) (LESS 5).
 GN RBGS-2A.
 OS Lycopersicon esculentum (Tomato).
 CC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta;
 CC eumhyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 CC core eudicots: Asteridae: euasterids I: Solanales; Solanaceae;
 CC Solanum.
 CC [1]
 RN SEQUENCE FROM N.A. (LESS 5).
 RC STRAIN-CV: YP36.
 RX MEDLINE: 87163513.
 RA MCNIGHT T.D., ALEXANDER D.C., BISCOCK M.S., SIMPSON R.B.:
 RT "Nucleotide sequence and molecular evolution of two tomato genes
 RT encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase.";
 RT Gene 48:23-32(1986).
 RL [2]
 RN SEQUENCE FROM N.A. (RBGS-2A).
 RX MEDLINE: 86233336.
 RA PICHERSKY E., BERNATZKY R., TANKSLEY S.D., CASHMORE A.R.:
 RT "Evidence for selection as a mechanism in the concerted evolution of
 RT Lycopersicon esculentum (tomato) genes encoding the small subunit of
 RT ribulose-1,5-bisphosphate carboxylase/oxygenase.";
 RT Proc. Natl. Acad. Sci. U.S.A. 83:3880-3884(1986).
 RL [3]
 RN SEQUENCE FROM N.A. (RBGS-2).
 RC STRAIN-CV: VENT CHERRY LA1221;
 RX MEDLINE: 88038372.
 RA SUGITA M., MANZARA T., PICHERSKY E., CASHMORE A., GRUSSEM W.:
 RT "Genomic organization, sequence analysis and expression of all five
 RT genes encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase from tomato.";
 RT Mol. Gen. Genet. 209:247-256(1987).
 RL [4]
 RN REVISIONS.
 RA MANZARA T.:
 RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP SEQUENCE OF 1-9 FROM N.A. (RBGS-2).
 RC STRAIN-CV: VENT CHERRY LA1221; TISSUE-ROOT;
 RX MEDLINE: 93144693.
 RA MANZARA T., CARRASCO P., GRUSSEM W.:
 RT "Developmental and organ-specific changes in DNA-protein interactions
 RT in the tomato rbcs1, rbcs2 and rbcs3a promoter regions.";
 RT Plant Mol. Biol. 21:69-88(1993).
 RL -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYLICOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M15236; AAA34192.1;
 CC DR EMBL: M1543; AAA34189.1;
 CC DR EMBL: X05983; CAA29401.2;
 CC

DR EMBL: X6069; CAA46869.1;
 DR PIR: S02363; RKT02.
 DR HSSP: P00866; IRLC.
 DR MENDEL: 414; LYCES:rbcs12.
 DR PFM: PFO0101; RUBISCO_small1:1
 KW Photosynthesis: Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 KW Maltose gene family.
 FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 58 180 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT VARIANT 87 87 I -> V (IN REF. 1).
 FT SEQUENCE 180 AA: 6BE7042D CRC32:
 RS2:LYCES Length: 180 February 14, 2000 08:02 Type: P Check: 4535 ..
 1 MASSIVISSA VAINSYTOA SWAPFTGLK SSATPYTKR QNIDITSIAS
 51 NGRVSCMOV WPIPMKRYE TSLYPLDSD EQLSEIEYL LKNGWPCLE
 101 FETEGFYVR ENKSPGYD GRWTWMLP MFGCTDATOV LAEVCARKA
 151 YPOAWRIIG FDNROYOCI SFIAVKPEGY
 !!NA: SEQUENCE 1.0
 ID RBGS3_SOLTU STANDARD; PRT; 181 AA.
 AC P32764;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN 3 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 3).
 GN RBGS-3.
 OS Solanum tuberosum (Potato).
 CC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta;
 CC eumhyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;
 CC core eudicots: Asteridae: euasterids I; Solanales; Solanaceae;
 CC Solanum.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV: AM 80.5793;
 RX MEDLINE: 94131296.
 RA FRITZ C.C., WOLTER F.P., SCHENKMEYER V., HERGET T., SCHREIER P.H.:
 RT "The gene family encoding the ribulose-(1,5)-bisphosphate
 RT carboxylase/oxygenase (Rubisco) small subunit of potato.";
 RT Gene 137:271-274(1993).
 RL -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYLICOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X69763; CAA49417.1;
 CC DR PIR: S31498; S31498.
 CC DR HSSP: P00866; IRLC.
 CC DR MENDEL: 495; SOLTU:rbcs3.6.
 CC PFM: PFO0101; RUBISCO_small1:1.
 CC

KW Photosynthesis: Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 FT Multiene family. 58 CHLOROPLAST (BY SIMILARITY).
 FT TRANSIT 1 58
 FT CHAIN 59 181 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT CHAIN 3.
 FT SEQUENCE 181 AA; 20437 MW; 268DC949 CRC32;
 RBS3_SCHPO Length: 181 February 14, 2000 08:02 Type: P Check: 4433 ..
 1 MASSIVSSAA VATRNSVQAQ SNAEPFTGLK SNAEPVTKK NNNVDTSIA
 51 SNGGRVCMQ WVPINMKKY ETLSTLPDL DEQLKEVEY LKNGWVPL
 101 EPEFHGFEVY RENHSPGY DGRYWTMNL PMFQCTDQ VLAEOEAKK
 151 AYPQAMIRLI GFDVROVOC ISFIAYKPEG Y
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 318 AA.
 ID RBSK_SCHPO
 AC 060116;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE RIBOKINASE (EC 2.7.1.15).
 GN SPC1655.02C.
 OS Schizosaccharomyces pombe (Fission yeast)
 OS Eukaryota; Fungi; Ascomycota; Ascomycetes;
 OS Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA LYNE M., WOOD V., RAJANDREAM M.A., BARRELL B.G., BECK A.,
 RA REINHARDT R.;
 RA Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + D-RIBOSE -> ADP + D-RIBOSE 5-PHOSPHATE.
 CC -1- PATHWAY: FIRST STEP IN RIBOSE METABOLISM.
 CC -1- SIMILARITY: BELONGS TO THE PERK FAMILY OF CARBOHYDRATE KINASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL023554; CAI19022.1;
 DR PROSITE: PS00583; PERK_KINASES_1; FALSE_NEG.
 DR PROSITE: PS00584; PERK_KINASES_2; 1.
 DR PIR: P00294; PERK; 1
 KW Hypothetical protein; Transferase; Kinase
 FT SEQUENCE 318 AA; 33250 MW; D40648BF CRC32;
 RBSK_SCHPO Length: 318 February 14, 2000 08:02 Type: P Check: 3214 ..
 1 MINIVLGSMTDLVMTKTI CPBGGETIIG EPDFSTGNG GRGANOAVAV
 51 ARLSNPADTK VSMGCVGD AFGVEMLSL KKGVDVNDV KTIENKSTGV
 101 AMIIVEETGE NRILLSEGAN GNDTAFVRA MEORISTCNL LIMELEIPLE
 151 AVEIALQIAH KHGVDVLMNP APAIPLSHDM ISYCAVLVFN EHEAAILNG
 201 ADSPATLENV DAYASKILSE GYKAVIITL GSQGYKYSA NGESALVSAC
 251 KYKAVDTAA GDTFGAFSN SIHGOPLKD SLEFAKCSA ITVQRGAAS
 301 SIPSLEVDG SFNLKNT
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 180 AA.
 ID RBS_MUSAC

AC 024045;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN PRECURSOR (EC 4.1.1.39)
 DE (RUBISCO SMALL SUBUNIT).
 GN RBCS1.
 OS Musa acuminata (Banana)
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphylliphytes; Spermatophyta; Magnoliophyta; Liliopsida;
 OC Zingiberales; Musaceae; Musa.
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE-LEAF;
 RA REGEV I., KHAYAT E., GERSTEIN S.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2.3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF008214; BAB63287.1;
 DR PIR: P01011; RUBISCO.Small; 1.
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide.
 FT TRANSIT 1 58
 FT CHAIN 59 180 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT CHAIN.
 FT SEQUENCE 180 AA; 20511 MW; 3BB9BCBD CRC32;
 RBS_MUSAC Length: 180 February 14, 2000 08:02 Type: P Check: 2518 ..
 1 MVSSMWTSSA AFTTRASPQ SSWAPFTGL KSAFAFVTR KPNADLSHP
 51 SNGGRVCMQ WVPINMKKY ETLSTLPDL DEQLKEVEY LKNGWVPL
 101 EPEFHGFEVY RENHSPGY DGRYWTMNL PMFQCTDQ VLAEOEAKK
 151 AYPQAMIRLI GFDVROVOC ISFIAYKPEG Y
 !!AA_SEQUENCE 1.0 STANDARD; PRT; 298 AA.
 ID REGS_DROME
 AC 094913;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHYTHMICALLY EXPRESSED GENE 5 PROTEIN (DREG-5).
 GN REG-5.
 OS Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S; TISSUE-HEAD;
 RX MEDLINE; 96203080.

RA VAN GELDER R.N., KRASNOM M.A.;
 RT "A novel circadianly expressed *Drosophila* melanogaster gene dependent
 RT on the period gene for its rhythmic expression.";
 RL EMBL J. 15:1625-1631(1996).
 CC -1- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS
 CC (POTENTIAL). IN THE HEAD, OSCILLATES IN ABUNDANCE WITH A DAILY
 CC PEAK DURING EARLY NIGHT, EVEN UNDER CONSTANT DARKNESS. OSCILLATION
 CC IS DEPENDENT ON PERIOD (PER) FUNCTION.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEAD, BUT NOT IN THE BODY.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN 24 HOURS EMBRYO.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 065103; AAC47267.1; -
 CC DR EMBLBASE: FB90015801; Reg-5.
 CC KW Biological rhythms.
 CC FT DOMAIN 22 25 POLY-SER.
 CC SQ SEQUENCE 298 AA; 32202 MW; F901A2C6 CRC32;
 REGS_DROME Length: 298 February 14, 2000 08:02 Type: P Check: 7903 ..
 1 MTTAKVILA CCLGAFHIQ ISSSSAIPIM EFLTRNEKMS HLYSTFAQV
 51 SVHCKSTAAV GLEPVNCKH NLIGYSARKL QTLSDVLEA LDPYORDANE
 101 LIWSSIMSDH PGASLVYTR OPLQOPLPTP PASSLILTR QQLPGASHA
 151 HPQSSGAT NPFESEGEK HKYAMDMDKA YGYQSSSE LVAALALSE
 201 PSKRRLTGPL VIRVRPDGSP VEEDKMPLP RDEDLPLTSS NGRSNAQA
 251 PQRNRNOLK AALRLHPAE RPPATPPDAE ASVPATGVR SRSDEPOA
 11AA_SEQUENCE 1.0
 ID REHY_TORRU STANDARD; PRT; 218 AA.
 AC P52574;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 13-DEC-1998 (Rel. 37, Last annotation update)
 DE REHYDRIN.
 OS Tortula ruralis (Moss).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;
 CC Bryidae; Pottiaceae; Tortula.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-STAR MOSS; TISSUE-LEAF;
 CC OLIVER M.J., SCOTT H.B. II;
 CC "Desiccation-tolerance and gene expression: Analysis of a recovery
 CC clone, Tr288, and its implications in mRNA storage during drying.";
 CC J. Exp. Bot. 45:577-583(1994).
 CC -1- FUNCTION: ASSOCIATED WITH THE REHYDRATION EVENTS INVOLVED IN THE
 CC RECOVERY OF THE DESICCATION-TOLERANT MOSS.
 CC -1- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. REHYDRIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 040818; AAA83758.1; -
 CC DR HSSP; P30041; IPRX.
 CC DR PFAM; PF00578; AHPC-TSA; 1.
 CC KW Antioxidant.
 CC SQ SEQUENCE 218 AA; 24084 MW; ADFD8F42 CRC32;

REHY_TORRU Length: 218 February 14, 2000 08:02 Type: P Check: 4850 ..
 1 MGGWALGDL VPDQADSTM GHKVDYCK DQWITFSHP GDVPPVCTTE
 51 LGKIAVNEP PEKRVKLLG LSTDIYEDNQ GMINDESYT PDAVLYPLIL
 101 ADDRKRTYA LNMDDPDKD ANKPPLASRA LHIIPDCRL KSLILPGTT
 151 GRNPDVLRV LDSLQASKH KATPANWCK GEPVYISPV SDEKANOMFP
 201 QGWETYNLRK ALRMTFVD
 11AA_SEQUENCE 1.0
 ID REP2_YEAST STANDARD; PRT; 296 AA.
 AC P03872;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE TRANS-ACTING FACTOR C (REP2) (PROTEIN CHARLIE).
 GN REP2.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Plasmid 2-micron.
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-A364A D5;
 CC MEDLINE: 81012161.
 CC RA HARTLEY J.L., DOWNSON J.F.;
 CC "Nucleotide sequence of the yeast plasmid.";
 CC Nature 286:860-864(1980).
 CC -1- FUNCTION: PLASMID PARTITION REQUIRE REP1, REP2, AND A CIS-ACTING
 CC DNA SEQUENCE (KNOWN AS STB).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J01347; AAB59343.1; -
 CC DR EMBL; V01323; CAA24633.1; -
 CC DR PIR; A04504; PDBYC.
 CC DR SGD; L0001612; REP2.
 CC KW Plasmid; Trans-acting factor.
 CC SQ SEQUENCE 296 AA; 33196 MW; E280A192 CRC32;
 REP2_YEAST Length: 296 February 14, 2000 08:02 Type: P Check: 6596 ..
 1 MODIEFANLU TYKARTAYV KDVCFLFTFM IAPVDIDIE SKRKDELLF
 51 PBYVIRPMES LITGRPGLD SAAEDSSVSS DSAEVLPLA AAMVKEKRS
 101 TENGALSSOE ASQAALDML QNNKLIDNRK QLYKSIATII GLPEKDKR
 151 ATEMLRKMD CTQLLVPPAP TEEDVAKLVV VTQLLTIVP PDROALIGD
 201 LFIPESLKI FNSFNELAE NQLQKSEL EGRTEVNHAN TNEVPSRRT
 251 RSRDINARGA YKLQNTTTEG PRAVPTKRR VATRVGRKS RNTSHV
 11AA_SEQUENCE 1.0
 ID REFB_NEIGO STANDARD; PRT; 346 AA.
 AC P37761;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE DTP-GLUCOSE 4,6-DEHYDRATASE (EC 4.2.1.46).
 GN REFB.
 OS Neisseria gonorrhoeae.

```

CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
CC [1]
CC SEQUENCE FROM N.A.
CC R. SERRAINE-MILL.
CC MEDLINE: 95050260.
CC ROBERTSON B.D., FROSCH M., VAN PUTTEN J.P.M.:
CC "The identification of cryptic rhamnose biosynthesis genes in
CC Neisseria gonorrhoeae and their relationship to lipopolysaccharide
CC biosynthesis." 176:6915-6920(1994).
CC J. Bacteriol. 176:6915-6920(1994).
CC -1- CATALYTIC ACTIVITY: DTDp-GLUCOSE = DTDp-4-DEHYDRO-6-DEOXY-D-
CC GLUCOSE + H(2)O.
CC -1- COFACTOR: NAD.
CC -1- PATHWAY: DTDp-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOLYSACCHARIDE BIOSYNTHESIS.
CC -1- SIMILARITY: STRONG, TO OTHER DTDp-GLUCOSE 4,6-DEHYDRATASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z33742; CAAB3652.1; -
CC EMBL: Z21508; CAAT9718.1; -
CC PIR: S47045; S47045.
CC PRF: PFO1370; EPImerase; 1.
CC K. Lipopolysaccharide biosynthesis; lyase; NAD.
CC NP BIND 13 19
CC SEQUENCE 346 AA; 38828 MW; FB9FD6A3 CRC32;
CC -----
REB_NEB Length: 346 February 14, 2000 08:02 Type: P Check: 5086 ..

1 MOTEKKNIL VTGGAGFICS AVYRHIIONT RDSYVINDLK TYAGNLESIT
51 DIADNPYAF EYVDICDRAE IDRYFAOYRP DAVMHLAES HYDRAIGSAG
101 EFIRNTIVGT FDLLEAARAY WOOMPSEKRE AFREHNISTD EYVGLHGTD
151 DLETTETPYA PSSPYASAKA AADHLVRAWQ RYRLSPISVS NCSNNGYPRQ
201 FPEKILPLMI LNALSGKPLP VYDGAQDIR WLEVEDHARA LYQVTEGVY
251 GETYVIGGN EKTINLEVVKI ICALLEELAP EKPAVAREY DLTTEVQDRP
301 GHDAVAYDA AKIRDLGWL PLETFESGR KTYQWYLDNK TRQNA

11AA_SEQUENCE 1.0
ID_RHE_MACMU STANDARD; PRT; 416 AA.
AC Q28849; 1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE RH-LIKE PROTEIN (RHESUS-LIKE PROTEIN).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Macaca.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-BONE MARROW;
RX MEDLINE: 94223693.
RA MOURO I., LE VAN KIM C., CHERIF-ZAHAR B., SALVIGNOL I., BLANCHER A.,
RA CARTON J.-P., COLIN Y.;
RA "Molecular characterization of the Rh-like locus and gene transcripts
RA from the rhesus monkey (Macaca mulatta).";
RA J. Mol. Evol. 38:169-176(1994).
CC -1- FUNCTION: MAY BE PART OF AN OLIGOMERIC COMPLEX WHICH IS LIKELY TO
CC HAVE A TRANSPORT OR CHANNEL FUNCTION IN THE ERYTHROCYTE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE RH FAMILY.

```

RT "Complete genomic sequence and analysis of 117 kb of human DNA
RT containing the gene BRCA1."
RL Genome Res. 6:1029-1049(1996).
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X85456; CAA64726.1; -
DR EMBL: L78833; AAC37595.1; -
DR HSSP: P06749; IABZ.
DR MIM: 601555; -
DR PFAM: PF00071; ras; 1.
DR GTP-binding; Prenylation; Lipoprotein.
KW NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 61 65 GTP (BY SIMILARITY).
FT NP_BIND 119 122 GTP (BY SIMILARITY).
FT DOMAIN 36 44 EFFECTOR REGION (POTENTIAL).
FT LIPID 224 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 227 AA; 25369 MW; 98964266 CRC32;
RH07_HUMAN Length: 227 February 14, 2000 08:02 Type: P Check: 2901 ..
1 MEGOSGRCKI VVGDAECGK TALQVFAKD AVPGSVPTV FENTASFEI
51 DKRIELNMW DTSGSSYYDN VRPLAYPDS AVLCFDSR PETIDSLVK
101 WQSEIQEFC NAKVYVGC LDMRTDLAT RELSQRLLP VTHQGVIVA
151 KQVANSIYE CSSRSERSY RQVHVATVA SLGRHQLR RTDSRBMQR
201 SAQLSGRPDR GNEGIHMDR AKSCNLM
11AA_SEQUENCE 1.0
RIB7_ARCFU STANDARD; PRT; 219 AA.
AC 028272;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PUTATIVE 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL REDUCTASE
DE (EC 1.1.1.193) (HTP REDUCTASE).
GN AF2007.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
CC [1]
RN SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KLENN K.A., DOSON R.J., GWINN M., HICKER E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERVANGS A.R., GRAM M.E., KAPRIDS N.C.,
RA FLEISCHMAN R.D., QUNACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEILL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA VASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL +
CC NAD(+) -> 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL + NADPH.
CC -1- PATHWAY: THIRD STEP OF RIBOFLAVIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE HTP REDUCTASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE00064; AAB89247.1; -
DR EMBL: AF2007; -
DR TIGR: AF2007; -
KW Hypothetical protein; Riboflavin biosynthesis; Oxidoreductase; NADP.
SQ SEQUENCE 219 AA; 24578 MW; 7DAEA807 CRC32;
RIB7_ARCFU Length: 219 February 14, 2000 08:02 Type: P Check: 4460 ..
1 NRPYEVNVA ASDGRISDE SRKQLRISCE EDLRIVDRIR AESDAIMVGI
51 GTVLADPRL TVKSAELREK RQDKGEENP LRVVDSRCR VPLTARIIND
101 EARTLVASR IAPERYREV KVAEVAFG EERVELSALL EFLHKKGVAR
151 LMVEGGGLI SSLISQNVLD EIRIYGPFI IGRDSDPTVC DQESFLKCR
201 IEKIRIGEG FAVTAFNR
11AA_SEQUENCE 1.0
RIB7_ARCFU STANDARD; PRT; 277 AA.
AC 000465;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE RIBOSOME-INACTIVATING PROTEIN LUFRIN ALPHA PRECURSOR (RNA
DE N-GLYCOSIDASE) (EC 3.2.2.22).
OS Luffa cylindrica (Smooch loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphyllophytes; Spermatophyta; Magnoliophyta; Eudicotyledons;
OC Core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
OC Luffa.
GN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-SEED;
RX MEDLINE: 92288316.
RA KATAOKA J., HABUKA N., MIYANO M., MASUTA C., KOIMAI A.;
RT "Nucleotide sequence of cDNA encoding alpha-luffin, a ribosome-
RT inactivating protein from Luffa cylindrica."
RL Plant Mol. Biol. 18:1199-1202(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RNA.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63371; CAA44229.1; -
DR PIR: S22494; S22494.
DR HSSP: P16094; IABZ.
DR PROSITE: PS00275; SHIGA_RICIN; 1.
DR PFAM: PF00161; RIP; 1.
KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin; Signal.
FT SIGNAL 1 19
FT CHAIN 20 277
SQ ACT SITE 179 179 RIBOSOME-INACTIVATING PROTEIN LUFRIN-
SQ SEQUENCE 277 AA; 30212 MW; E0233100 CRC32;
RIP7_LUCFY Length: 277 February 14, 2000 08:02 Type: P Check: 429 ..

```

1 MKRFTVLLIA IFVAATVEA DVRESLGSS STSYSKETID LRKALPENGCT
51 VYNITLLSS ASGASRYTLM TLSNYDGKAI TAAVDVTANY IMGVLVNSTS
101 YFENEDAKL ASQYFKSGT IVTLPSGNY EKLQTAAGRI REKIPLGEP
151 LDSAITTLFH YDSTAAAF LVLTQTAEA SRFKIEGDI IERISKNOVP
201 SLATILENE WSALSKOJOL AQTNNGTFRK PYVIIDDKQ RVELTNTSK
251 VYTKNTOLL NKONVAAP EDVSARK

11AA_SEQUENCE 1.0
ID RPB_LUFCY STANDARD: PRT: 250 AA.
AC P24613;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE RIBOSOME-INACTIVATING PROTEIN LUFIN-B (RRNA N-GLYCOSIDASE)
DE (BC 3.2.2.22).
OS Lufta cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids I; Cucurbitales; Cucurbitaceae;
OC Lufta.
RN [1]
RP SEQUENCE.
RC TISSUE-SEED;
RE MEDLINE: 91248488.
RA ISLAM M.R., HIRAYAMA H., FUNATSU G.;
RT "Complete amino acid sequence of luftin-b, a ribosome-inactivating
RT protein from sponge gourd (Lufta cylindrica) seeds."
RL Agric. Biol. Chem. 55:229-238(1991).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC AENOSINE ON THE 28S RRNA.
CC -1- SIMILARITY: NO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS BELONGS TO TYPE 1 RIP.
DR PIR: J01008; J01008.
DR HSP: P16094; IMOM.
DR PROSITE: PS00275; SHIGA_RICIN; 1.
DR PFAM: PF00151; RIP; 1.
KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin.
FT ACT_SITE 160 160 BY SIMILARITY.
SQ SEQUENCE 250 AA; 27293 MW; 6A472993 CRC32;

RIPB_LUFCY Length: 250 February 14, 2000 08:02 Type: P Check: 7681 ..

1 ANVSFSLSGA DSKSYSKFT ALRKALPSKE KVSNIPLILP SASGASRYIL
51 MOLSNDYDAA ITMAIDVTNY YIMGVLVNST SYFANESDAR LASQYFKGS
101 TVLTIPYSGN YERLONAGK IREKIPLGFR ALDSALTSIF HYDSTAAAF
151 FLVILQTTAE ASRFKIEGDI IERIKPNEV PSPALSLBN EAWSLSKQI
201 OLMOITNGAF RTPVIIDNK GORVEITNLA SKQOINDVNS KLLNKONIA

11AA_SEQUENCE 1.0
ID R21_SPIOL STANDARD: PRT: 256 AA.
AC P24613;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 30S RIBOSOMAL PROTEIN L21, CHLOROPLAST PRECURSOR (CL21) (CS-L7).
GN RPL21.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Caryophyllidae; Caryophyllales; Chenopodiaceae;
OC Spinacia.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 56-67.
RE MEDLINE: 91168304.

RA MARTIN W., LAGRANGE T., LI Y.F., BISANT-SEYER C., MACHE R.;
RT "Hypothesis for the evolutionary origin of the chloroplast ribosomal
RT protein L21 of spinach."
RL Curr. Genet. 18:553-556(1990).
RN [2]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 56-75.
RC STRAIN-CV. ALVARO;
RE MEDLINE: 90375547.
RA SMOOKER P.M., KRUEF V., SUBRAMANIAN A.R.;
RT "A ribosomal protein is encoded in the chloroplast DNA in a lower
RT plant but in the nucleus in angiosperms: isolation of the spinach L21
RT protein and cDNA clone with transit and an unusual repeat sequence."
RL J. Biol. Chem. 265:16699-16703(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. GEANT D'HIVER;
RE MEDLINE: 93205007.
RA LAGRANGE T., FRANZETTI B., AXELOS M., MACHE R., LERBS-MACHE S.;
RT "Structure and expression of the nuclear gene coding for the
RT chloroplast ribosomal protein L21: developmental regulation of a
RT housekeeping gene by alternative promoters."
RL Mol. Cell. Biol. 13:2614-2622(1993).
CC -1- FUNCTION: THIS PROTEIN BINDS TO 23S RIBOSOMAL RNA IN THE PRESENCE
CC OF PROTEIN L20 (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L21P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56691; CAA40019.1; -
DR EMBL: M57413; AAA34041.1; -
DR EMBL: M64682; AAA74715.1; -
DR PIR: S13527; S13527.
DR PIR: A48103; A48103.
DR MENDEL: 10833; SPIOL; rpl21.1.
DR PROSITE: PS01169; RIBOSOMAL_L21P; 1.
DR PFAM: PF00829; RIBOSOMAL_L21P; 1.
KW Ribosomal protein; rRNA-binding; Chloroplast; Transit peptide.
FT TRANSIT 1 55 CHLOROPLAST.
FT CHAIN 56 256 50S RIBOSOMAL PROTEIN L21.
FT DOMAIN 236 253 ALA/GLU-RICH.
FT SEQUENCE 256 AA; 28408 MW; D1FB7F00 CRC32;

RK21_SPIOL Length: 256 February 14, 2000 08:02 Type: P Check: 718 ..

1 MASATLAFSC SSLCATLKL P ONLNPILLNV PLSKPFEGV VSPPSLSRLS
51 LIPVAKRRR FOEIPBELKA EFFEFORPPN QKPQLSDVLP DDFQAREPGT
101 PEYNDINGOF LPKKGPPPR EEIFANVVG SROQVIVGR WIYQRLKGA
151 TVNDKIVLNR VLVGTAKST YIGTPIVTA AVHANVEQL LDDKIVFKY
201 KKKKNRRNI GHROPITRIK ITGITGYEDY PASTLEAVE AKKEAAEAENE
251 AEAVPY

11AA_SEQUENCE 1.0
ID R8_MARPO STANDARD: PRT: 132 AA.
AC P06362;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S8.
GN RPS8.
OS Marchantia polymorpha (Liverwort).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;
OC Chloroplast.

```

OC Marchantiales; Marchantiaceae; Marchantia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 8906687.
RA FUKUZAWA H., KOHCHI T., SANO T., SHIRAI H., UMESONO K., INOKUCHI H.,
RA OZERI H., OHYAMA K.,
RT Structure and organization of Marchantia polymorpha chloroplast
genome. III. Gene organization of the large single copy region from
rt rbcL to trnI (CAU).
RT J. Mol. Biol. 203:333-351(1988).
RN [2]
RP COMPLETE GENOME.
RA OHYAMA K., FUKUZAWA H., KOHCHI T., SHIRAI H., SANO T., SANO S.,
RA UMESONO K., SHIKI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,
RT Chloroplast gene organization deduced from complete sequence of
rt liverwort Marchantia polymorpha chloroplast DNA.
RL Nature 322:572-574(1986).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.
CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X04465; CAA28121.1; .
DR PIR: A02716; R3LV8.
DR HSSP: P56209; 1SET.
DR MENDEL: 13155; MARPO:ps8.1.
DR PROSITE: PS00053; RIBOSOMAL_S8: 1.
DR PFAM: PF00410; Ribosomal_S8: 1.
KW Ribosomal protein; rRNA-binding; Chloroplast.
SO SEQUENCE 132 AA; 14921 MW; 20331A62 CRC32;
RR_MARPO Length: 132 February 14, 2000 08:02 Type: P Check: 6002 ..
1 MGNDTIANMI TSIRNANLKG IKTVQYPAIN ITRNIATKLF QEGFIDNFID
51 NKONTKDILI LNKYQGRK KSYITTLRI SKPLRIYSN HKELPKVLGG
101 MGIVILSTR GIMTDREARQ KRIGELLGY VW
11AA_SEQUENCE 1.0
ID RS13_SCHPO STANDARD; PRT; 150 AA.
AC P28189;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 40S RIBOSOMAL PROTEIN S13.
GN RPS13 OR SPAC6F6.07C.
OS Eukaryotes; Archaea; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92375702.
RA MARKS J., SIMANIS V.,
RT Cloning of the gene for ribosomal protein S13 from the fission yeast
rt Schizosaccharomyces pombe.
RL Nucleic Acids Res. 20:4094-4094(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA GENTILES S., CHURCHER C.M., BARRELL B.G., RAJANDREAN M.A., WOOD V.,
RL submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X67030; CAA47424.1; .
DR EMBL: Z98981; CAB1741.1; .
DR PIR: S26296; S26296.
DR PROSITE: PS00362; RIBOSOMAL_S15: 1.
DR PFAM: PF00312; Ribosomal_S15: 1.
KW Ribosomal protein.
KW INT.MET
FT SEQUENCE 150 AA; 16822 MW; C7588E0E CRC32;
RS13_SCHPO Length: 150 February 14, 2000 08:02 Type: P Check: 3858 ..
1 GRMSKKGKI ASSALPYRS PPAACKADAD SYVEQLKPS KKGSPSOIG
51 VTLDSHGIP QVREITGQKI MRILKANGLA PELPELVYL IKKAVYARKH
101 LERNRKQKDS KFFLLIESR IHLARYRK VGALPPTWKY ESATASALVA
11AA_SEQUENCE 1.0
ID RS3_ACHAX STANDARD; PRT; 257 AA.
AC P41117;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S3.
GN RPS3 OR RPS3.
OS Achleplasma axanthum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Achleplasmataceae; Achleplasma.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-5743.
RX MEDLINE: 94169035.
RA TOTH K.T., HARRISON N., SEARS B.B.,
RT Phylogenetic relationships among members of the class Mollicutes
rt deduced from rps3 gene sequences.
RL Int. J. Syst. Bacteriol. 44:119-124(1994).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE BINDING OF INITIATOR
CC MET-tRNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L22465; AAA21916.1; .
DR PROSITE: PS00548; RIBOSOMAL_S3: 1.
DR PFAM: PF00189; Ribosomal_S3_C: 1.
DR PFAM: PF00417; Ribosomal_S3_N: 1.
KW Ribosomal protein; tRNA-binding.
KW SEQUENCE 257 AA; 28717 MW; 8035520F CRC32;
RS3_ACHAX Length: 257 February 14, 2000 08:02 Type: P Check: 7310 ..
1 MGRVNPDIGF RVGVIRDMDS KMYADKRIIP ALVKEDAVIR KFLNKKYRNA
51 AVSHEIERL KELVKRRVK ITLHGKPGV VIGREAYTRK ETIASLEKLT
101 KKEIVFWVE VRKPEVATL VAQSMAROLE NRASFESKTR NCYAKSIKVR
151 AKGIKTSOR LGREWARTE GYSEGQVPLH TLRADEVAT AEAOTYGIL
201 GIKWITHE ILPGQSHETL RKEOSSASS NNGGGRARRS RKGPRRSGED

```

251 NATEGNN
11AA_SEQUENCE 1.0
ID RS8_SYNY3 STANDARD: PRT: 133 AA.
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30S RIBOSOMAL PROTEIN S8.
GN RPS8 OR RPS8 OR SL1809.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUDO K.,
RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YAMADA M., YASUDA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RT DNA Res. 9:109-136(1996)
RL -1- FUNCTION: BINDS DIRECTLY TO THE CENTRAL DOMAIN OF 16S RIBOSOMAL
CC RNA (BT SIMILARITY). BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90905; BAA17336.1;
CC DR HSSP: P56209; 1SER.
CC DR PROSITE: PS00053; RIBOSOMAL_S8; 1.
CC DR PFM: PF00410; Ribosomal_s8; 1.
CC KW Ribosomal protein: rRNA-binding.
CC SEQUENCE 133 AA; 14666 MW; 3596E999 CRC32;
RS8_SYNY3 Length: 133 February 14, 2000 08:02 Type: P Check: 8926 ..

1 MASTDISDM LTRIRNCAV RSTQVPT KMTLSIAVL KSGFIEDYS
51 ETGEGINKML VTLKYGKT RQPLNTIQ VSKRLRYYS PSKRIRVGL
101 GIGIAIVTS HGIMTDEAR RGIGIGELIC YIV

11AA_SEQUENCE 1.0
ID RUVB_TRETH STANDARD: PRT: 324 AA.
AC 056214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE HOLLIDAY JUNCTION DNA HELICASE RUVB.
GN RUVB.
OS Thermus aquaticus (subsp. thermophilus).
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96196179.
RA TONG J., WETZEL J.G.;
RT "Cloning, sequencing, and expression of ruvb and characterization of
RT ruvb proteins from two distantly related thermophilic eubacteria.";
RT J. Bacteriol. 178:2695-2700(1996)
RL -1- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP REPAIRS
CC CRUPTURE STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE.
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC HOMOLOGOUS RECOMBINATION. RUVB IS AN HELICASE THAT MEDIATES THE

```

```

HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
REANNEALING (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A COMPLEX WITH RUVA.
CC -1- SIMILARITY: NO OTHER ATP-BINDING DNA REPAIR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U22817; AAB03726.1;
CC DR PFM: PF00004; AAA; 1.
CC DR DNA repair; SOS response; ATP-binding; DNA recombination; Helicase.
CC NP_BIND 45 52
CC FT SEQUENCE 324 AA; 36015 MW; D9638BD3 CRC32;
SQ SEQUENCE 324 AA; 36015 MW; D9638BD3 CRC32;
RUVB_TRETH Length: 324 February 14, 2000 08:02 Type: P Check: 3575 ..

1 MEDIALPRT LDEYGOERL KOKLRVYLEA AKARKEPLEH LLFGPGGLG
51 KTLAIVIAH ELGVNLRYTS GALEKPGDL AAILANLEE GDILFIDEIH
101 RLSQAQEHV YPAMEDVMD IVIGGPAAR TIRLELPRA LIGATRPGL
151 ITALLSRFG IVEHLEYTP ELAQGVMD ARLGVIRE EALGIGRS
201 RGTNRVAKRL FRVRDFAY EGEEVTRER ALEALAAOL DELLEKDR
251 EILEVLLRF GAGPGVGLAT ATALSEDPGT LEEVHEPILI RQGLKRTPR
301 GRVATELAIR HGYPEPVGP LLEP

11AA_SEQUENCE 1.0
ID S27A_ECOLI STANDARD: PRT: 217 AA.
AC P26428; P76673;
DT 01-NOV-1992 (Rel. 23, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SIGMA CROSS-REACTING PROTEIN 27A (SCP-27A).
GN YHBL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / W3110;
RA SMITLIE D.A., FUJITA N., TOWNSLEY F.M., ISHIIHAMA A., HAYWARD R.S.;
RL submitted (xxx-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA BLATTNER F.R., FLOURENT G., III, BLOCH C.A., PERRA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-24.
RX MEDLINE: 92246944.
RA UESHIMA R., FUJITA N., ISHIIHAMA A.;
RT "Identification of Escherichia coli proteins cross-reacting with
RT antibodies against region 2.2 peptide of RNA polymerase sigma
RT subunit 2."
CC Biochem. Biophys. Res. Commun. 184:634-639(1992).
CC -1- FUNCTION: NOT KNOWN. CROSS-REACTS WITH ANTIBODIES AGAINST REGION
CC 2.2 OF RPOD AND RPOH.
CC -1- SIMILARITY: BELONGS TO THE ES1 FAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D13188; AAA02487.1; -
DR EMBL: U18997; AAA58011.1; ALT_INT.
DR EMBL: A000400; AAC76241.1; ALT_INT.
DR PIR: JN0287; JN0287.
DR ECOGENE: EG11383; YHBL.
FT CONFLICT 111 111 C -> S (IN REF. 1).
FT CONFLICT 195 217 FPACWELNE (IN REF. 1).
SQ SEQUENCE 217 AA; 22981 MW; 501BEDA CRC32;

s37a_ECOLI Length: 217 February 14, 2000 08:02 Type: P Check: 54

1 MKRIGVILSG CGVYDSEIH EAVLTLLAIS RSGAAYCFA PDKOYDVIN
51 HLTGAMTET RNVLEIAR I TGEIRPLAQ ADAALDALI VPGFGAAN
101 LSNFASLGE CTVDRELKAL AQAMHAGKP LGFACIAPAM LPRIEDELAR
151 LTIGDIDTA EVLEMGAEH VPCPDIVY DEDKIVTTP AYVLAQNIAE
201 AASGIDKLVS RVLYAE

11AA_SEQUENCE 1.0 STANDARD; PRT; 433 AA.

AC SEQRES: 433
ID STRAIN: 15-DEC-1998 (Rel. 39, Created)
DT 15-DEC-1998 (Rel. 39, Last sequence update)
DT 15-DEC-1998 (Rel. 39, Last annotation update)
DE PREPROMETIN TRANSLOCASE SECY SUBUNIT.
GN SECY OR R639.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.

RA STRAIN-MADRID E.
RA MEDLINE: 99039499.
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SIEHERITZ-PONTEN T., ALSMARK U.C.M., PODOMSKI R.M., NAEISUND A.K.,
RA ERIKSSON A.-S., WINKLER H.H., KURLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).

CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECY AND SECY
CC TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
CC BY FORMING PART OF A CHANNEL. (BY SIMILARITY).
CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECY-F & SECY) THAT
CC COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SECY/SECY-ALPHA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ235272; CA15079.1; -
DR PROSITE: PS00755; SECY_1; FALSE_NEG.
DR PROSITE: PS00756; SECY_2; 1.
KW Protein transport; Transmembrane; Translocation.
FT TRANSMEM 17 37 POTENTIAL.

FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 268 288 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
SQ SEQUENCE 433 AA; 47696 MW; 2A88EC17 CRC32;

SECY_RICPR Length: 433 February 14, 2000 08:02 Type: P Check: 3620

1 MGONFESKSS NDVNRRIIT LFMILICRNG SEPIPIGIDS IALNSVAEKN
51 QFGILGMFNM LSGSLGHS IFALAIPIY TASIIQLMS VAYKPLENEK
101 KEGETGKRI NOLSRILTVL LASEQAYGVA LSESMVTNT GPVILLAGEF
151 FRVTIVITLV VGTILMLWG EQITORGIGN GTSILIFIGI ISGVPAIIS
201 MFLSRKAL SPLAIIVCI GAVLLAIITII FFEKQKRL VQYKROVGN
251 KIYGEATHM PLKNTSGVI PPIFASILL FPTLASFSN SNSDTSMILT
301 YLGHGKPYV ILKLVYLIME FSPFTAIIV NSEETANLR KYGAYIPGR
351 PKNTSDYFD YILRLIVIG GILSVICVY PELLMKRYVI SLSLGSSTEL
401 IVNVVLDTM TQIOTYLFSS KYEGLMKKIK LKN

11AA_SEQUENCE 1.0 STANDARD; PRT; 161 AA.

ID STRAIN: 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATASE ACTIVITY TOWARDS THE HPT DOMAIN OF THE ARCB SENSOR
DE PHOSPHOISTIDINE PHOSPHATASE SIXA (EC 3.1.3.-) (RX6).
GN SIXA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

RA STRAIN-K12 / MC4100.
RA MEDLINE: 98149313.
RA OGINO H., MATSUBARA M., KATO N., NAKAMURA Y., MIZUNO T.;
RA "An Escherichia coli protein that exhibits phosphatidyl
RT phosphatase activity towards the Hpt domain of the Arcb sensor
RT involved in the multistep His-Asp phosphorelay."
RL Mol. Microbiol. 27:573-585(1998).

CC -1- FUNCTION: EXHIBITS PHOSPHOISTIDINE PHOSPHATASE ACTIVITY TOWARDS
CC THE HPT DOMAIN OF THE ARCB SENSOR INVOLVED IN THE MULTISTEP HIS-

CC The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONDO A., HORIUCHI T.,
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KAWAI K., KASAI H.,
RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAGAWA M., MAKINO K.,
RA MASUDA S., MIKI T., MIYOSUCHI K., MORI H., MOTOKURA K., NAKAMURA Y.,
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SATO N., SAMPEI G., SEKI Y.,
RA TAGAMI H., TAKEKOTO K., WADA C., YAMAMOTO Y., YANO M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: EXHIBITS PHOSPHOISTIDINE PHOSPHATASE ACTIVITY TOWARDS
CC THE HPT DOMAIN OF THE ARCB SENSOR INVOLVED IN THE MULTISTEP HIS-


```

CC ASP PHOSPHORELAY.
CC -1- SIMILARITY: BELONGS TO THE SIXA FAMILY OF PHOSPHATASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D86398; BAA24878.1; -
DR EMBL: AB000322; AAC5400.1; -
DR EMBL: D90864; CAB22123.1; -
DR EMBL: D90865; CAB22129.1; -
DR EMBL: EGI4126; SIXA.
DR ECOCENE: EGI4126; SIXA.
KW Hydrolyase.
FT CONFLICT 80 80 T -> K (IN REF. 3).
SQ SEQUENCE 161 AA; 17208 MW; 396B6F03 CRC32;

SIXA_ECOLI Length: 161 February 14, 2000 08:02 Type: P Check: 9360

1 MGVFIMRHGD AALDAASDV RPLTNGCDE SRLMANMLKG QKVEIERVLV
51 SPFLAEQTL EEVGDCLNP SSAEVLPELT PCGDVGLVSA YLQALNTEGV
101 ASVLISHLP LVGYLVAELC PGETPMFTT SALASVTLDE SGNGTNMQM
151 SPCNKKMAKA I

11AA_SEQUENCE 1.0
ID SIXA_HAEIN STANDARD; PRT; 164 AA.
AC P44164.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHOTRISIDINE PHOSPHATASE SIXA HOMOLOG (EC 3.1.3.-).
GN SIXA OR H11338.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-RD / KW20;
RC MEDLINE: 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIPLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,
RA FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GENEW C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- SIMILARITY: BELONGS TO THE SIXA FAMILY OF PHOSPHATASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32813; AAC22985.1; -
DR TIGR: H11338; -
KW Hydrolyase.
SQ SEQUENCE 164 AA; 18700 MW; 3F6E5D03 CRC32;

SIXA_HAEIN Length: 164 February 14, 2000 08:02 Type: P Check: 2974

```

```

1 MNIFMERGE AEYMANSKRA RHLVYGSQK AFLQOGWLKO HSLTVLNSL
51 DRIVSPYVR ADETFHOVNO AFDELEENKF EIWEGITPYG HAHSVIDYLE
101 VLKDEGVKSV LIVSHPLVG EIVAEVYGRK NPISFYPATI AOLWOGNKS
151 EILMQASPV IYLK

11AA_SEQUENCE 1.0
ID SPAL_SALTY STANDARD; PRT; 432 AA.
AC P39444.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PROBABLE ATP SYNTHASE SPAL/INVC (EC 3.6.1.34).
GN SPAL OR INVC.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-SRLT / SL1344;
RC MEDLINE: 94321319.
RA EICHELEBERG K., GINOCCHIO C.C., GALAN J.E.;
RT "Molecular and functional characterization of the Salmoneilla
RT typhimurium invasion genes invB and invC: homology of invC to the
RT F0F1 ATPase family of proteins."
RT J. Bacteriol. 176:4501-4510(1994).
[2]
RN SEQUENCE OF 97-432 FROM N.A.
RP MEDLINE: 94008985.
RA GROISMAN E.A., OCHMAN H.;
RT "Cognate gene clusters govern invasion of host epithelial cells by
RT Salmoneilla typhimurium and Shigella flexneri."
RT EMBO J. 12:3779-3787(1993).
[3]
RN SEQUENCE OF 403-432 FROM N.A.
RP STRAIN-SRLT / SL1344;
RC MEDLINE: 95272391.
RA COLLAZO C., ZIERKER M.K., GALAN J.E.;
RT "Functional analysis of the Salmoneilla typhimurium invasion genes
RT invI and invJ and identification of a target of the protein secretion
RT apparatus encoded in the inv locus."
RT Mol. Microbiol. 15:25-38(1995).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT ENTRY OF S.TYPHIMURIUM INTO
CC CULTURED EPITHELIAL CELLS. PROBABLE CATALYTIC SUBUNIT OF A PROTEIN
CC TRANSLOCASE. MAY ENERGIZE THE PROTEIN EXPORT APPARATUS ENCODED IN
CC THE INV LOCUS WHICH IS REQUIRED FOR THE SURFACE PRESENTATION OF
CC DETERMINANTS NEEDED FOR THE ENTRY OF SALMONEILLA SPECIES INTO
CC MAMMALIAN CELLS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: TO OTHER TRANSPORT APPARATUS AND TO ATPASE
CC BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08279; AA474038.1; -
DR EMBL: X73525; CAAS1921.1; -
DR EMBL: U10872; AA83429.1; -
DR PIR: S37304; S37304.
DR STGENE: SG10464; SPAL.
DR POSITE: PS00152; ATPASE_ALPHA_BETA. 1.
DR PFAM: PF00006; ATP-synt_ab. 1.
KW Hydrolyase; Hydrogen ion transport; ATP synthesis; ATP-binding;
KW virulence.
SQ NP_BIND 159 166 ATP (POTENTIAL).
FT

```

FT MUTAGEN 165 165 K->E: LOSS OF FUNCTION.
 FT CONFLICT 291 291 G->A (IN REF. 2).
 FT CONFLICT 388 401 CARSTITIGRCRG -> GENIDNBRAMQR (IN
 REF. 2)
 FT CONFLICT 407 407 R->W (IN REF. 2).
 FT CONFLICT 432 AA: 47253 MW: DC79488F CRC32:
 SPAL_SALTY Length: 432 February 14, 2000 08:02 Type: P Check: 7477

1 MKTRLLQYL AYPQKITGPI IEAELRDYAI GELCEITPWL AKRTGCCTCA
 51 GGMVLATGTHR ADAYRNCQGL SBDVLYPTG RALSAMWYGS VIGAVLDPTG
 101 KIVERFTPEV APISEERVID VAPPSYASRV GVREPLITGV RAIDGLITCG
 151 VGORMGIFAS AGCGKTMLMH MLEQTEADV FVIGLIGERG REVTEFVDM
 201 RASHKKECV LVEFATSDFPS VDRCAAOIA TTVAEYFRQ GRYVLYFIDS
 251 MTRARALRD VALASERPA RRGYPASYVD NLPILLRPG GTSQSITAF
 301 YTVLESEEE ADPMADERS IIDGHLYSR KLAGQGHYPA IDVLSVSRV
 351 FGQYTTPTHA EQASAVRKLH TRLELQLEFI DLGEYRCAK ISITIGRCRC
 401 GDSUKARLCQ PVAQYSSFD TLSCMNAFAD QN

11AA_SEQUENCE 1.0 STANDARD; PRT: 241 AA.

ID SPIR_SPIME
 AC P21625: 18, Created
 DT 01-MAY-1991 (Rel. 18, last sequence update)
 DT 01-MAY-1991 (Rel. 18, last annotation update)
 DE SPIRALIN.
 OS Spiroplasma melliferum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Spiroplasmataceae; Spiroplasma.
 RN [1]

RC SEQUENCE FROM N.A.
 RA STRAIN-ATCC 33219 / BC3;
 RX MEDLINE: 91008990.
 RA CHEVALIER C., SAILLARD C., BOVE J.M.;
 RT "Spiralins of Spiroplasma citri and Spiroplasma melliferum: amino
 acid sequences and putative organization in the cell membrane.";
 RL J. Bacteriol. 172:6090-6097(1990).

CC -1- SUBUNIT: SEEMS TO OCCUR AS DIMER, TETRAMERS, AND LARGE OLIGOMERS
 OF IDENTICAL CHAINS.
 CC -1- SUBCELLULAR LOCATION: THIS IS THE MAJOR MEMBRANE PROTEIN OF
 SPIROPLASMA.

CC -1- PPM: ACYLATED BY A PALMITIC ACID GROUP ON A THR.
 CC -1- SIMILARITY: 75% TO THE S.CITRI SPIRALIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: M59366; AAA26588.1; -
 DR PIR: A36149; A36149.
 KW Transmembrane; Palmitate; Lipoprotein.
 FT TRANSMEMBRANE 0 PROBABILE.
 FT DOMAIN 1 164 EXTRACELLULAR.
 FT TRANSMEM 165 184 POTENTIAL.
 FT DOMAIN 185 241 CYTOPLASMIC.
 FT LIPID 170 170 PALMITATE (POTENTIAL).
 SO SEQUENCE 241 AA: 25310 MW: 2868DFD CRC32:

SPIR_SPIME Length: 241 February 14, 2000 08:02 Type: P Check: 9527
 1 KKLSTIAVF GVSAGVTTSY VACNKTESNN LSKYKTIAP ATVAASTRKA

51 VTKPEIKTL EANVLKAVOG VKTATADE QFEVYKNSKG TALETIDLEA
 101 GKVEVLIQIT PAKDKTVIG ETRIKVTLR KIGEVTKVDI KDVIYEQTV
 151 GIASSTPKAV KDELNAVNT YATLAKAVLD ALQNTAPNAG ASDPEITNNG
 201 AEGDYENAKE VEYTYKAND SANISGOFK KAKYATAPT E

11AA_SEQUENCE 1.0 STANDARD; PRT: 134 AA.

ID SRP_CHLPS
 AC P28164:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 01-OCT-1996 (Rel. 34, last annotation update)
 DE SULFUR-RICH PROTEIN.
 GN SRP.
 OS Chlamydia psittaci.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-6BC;
 RX MEDLINE: 91267949.
 RA EVERETT K.D.E., HATCH T.P.;
 RT "Sequence analysis and lipid modification of the cysteine-rich
 envelope proteins of Chlamydia psittaci 6BC.";
 RL J. Bacteriol. 173:3821-3830(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: M61116; AAB61620.1; -
 SO SEQUENCE 134 AA: 14732 MW: CC0422F5 CRC32:

SRP_CHLPS Length: 134 February 14, 2000 08:02 Type: P Check: 9648

1 MGENSNISIG SDVTSLIQPG LEQVODEGV QVSLINSVLG MCRVHINPI
 51 KTSKIVQSRRA FOITWVGLI ILLIAGLALT FVLQGLQGN AFLFLPAVI

101 GLVKLATSV CMERCTPER WRLCKRLMOQ LKIF

11AA_SEQUENCE 1.0 STANDARD; PRT: 352 AA.

ID SSAU_SALTY
 AC P96069:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, last sequence update)
 DT 15-DEC-1998 (Rel. 39, last annotation update)
 DE SECRETION SYSTEM APPARATUS SSAU.
 GN SSAU.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 97285756.
 RA HENSEL M., SHEA J.E., RAUPACH B., MONACK D., FALKOW S., GLEESON C.,
 RA KUBO T., HOLDEN D.W.;
 RT "Functional analysis of ssau and the ssak/v operon, 13 genes encoding
 components of the type III secretion apparatus of Salmonella
 pathogenicity island 2.";
 RL Mol. Microbiol. 24:155-167(1997).

CC -1- FUNCTION: PART OF A TYPE III SECRETION SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE FLHB/HRPV/YSCU/SPAS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: X99944: CAA68202.1: -
 DR STYGENE: SG372772: SSN: 1
 DR PFM: P01312: BacExport_2: 1
 DR Transport: Protein transport: Transmembrane.
 FT TRANSMEM 34
 FT TRANSMEM 89
 FT TRANSMEM 109
 FT TRANSMEM 144
 FT TRANSMEM 176
 FT TRANSMEM 196
 FT TRANSMEM 352 AA: 39760 MW: E7C787BA CRC32;
 SO SEQUENCE

SSAU_SALT Length: 352 February 14, 2000 08:02 Type: P Check: 3263

1 MSEKEQPT KLRDGRKG QVKSIEITS LFOILALYLY FHEFEKML
 51 ILISITFTL QLVNKPFSYA LTOJSHALIE SLTSALLFLG AGVIAVGS
 101 VFLQGVVIA SKAIGKSEH INPVSNFKOI FSLHVEYELC KSLKIVMLS
 151 LIFAFFYYKY ASFRALPYC GLAGVLVVS SLIKMLWGV MVEYIVGIL
 201 DYSQYKIR KDLKSKDGV KOEHKDEGD PQMKTRREM QSEIGSSLA
 251 QSVQSVAVV RNPTIAVCL GYHPDMP RYLEKSDAQ ANYIVNIAR
 301 NCIPVENVE LARSLEFEVE RGDKIPETLF EPVALLRV MKIDVHSTE
 351 TP

11AA_SEQUENCE 1.0 STANDARD: PRT: 391 AA.
 ID SSRL_HUMAN
 AC P30873:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE SOMATOSTATIN RECEPTOR TYPE 1 (SSIR) (SRIF-2).
 GN SSRI.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92108031.
 RA YAMADA Y., POST S.R., WANG K., TAGER H.S., BELL G.I., SEINO S.:
 RT "Cloning and functional characterization of a family of human and
 RT mouse somatostatin receptors expressed in brain, gastrointestinal
 RT tract, and kidney."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:251-255(1992).
 CC -1- FUNCTION: RECEPTOR FOR SOMATOSTATIN WITH HIGHER AFFINITY FOR
 CC SOMATOSTATIN-14 THAN -28. THIS RECEPTOR IS COUPLED VIA PERTUSSIS
 CC TOXIN SENSITIVE G PROTEINS TO INHIBITION OF ADENYLATE CYCLASE. IN
 CC ADDITION IT STIMULATES PHOSPHORYLASE PHOSPHATASE AND NA+/H+
 CC EXCHANGER VIA PERTUSSIS TOXIN INSENSITIVE G PROTEINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: FETAL KIDNEY, FETAL LIVER, AND ADULT PANCREAS,
 CC BRAIN, LUNG, JEJUNUM, AND STOMACH.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

EMBL: M81829: AAS58247.1: -

DR PIR: A41795: A41795.
 DR GCRDB: GCR_0272: -
 DR MIM: 182451: -
 DR PROSITE: P500237: G-PROTEIN_RECEPTOR. 1.
 DR PFM: P01001: 7cm.1: 1.
 DR G-protein coupled receptor: Transmembrane; Glycoprotein;
 KM Multigene family: Lipoprotein; Palmitate.
 FT TRANSMEM 1
 FT TRANSMEM 57
 FT TRANSMEM 84
 FT TRANSMEM 85
 FT TRANSMEM 95
 FT TRANSMEM 120
 FT TRANSMEM 121
 FT TRANSMEM 131
 FT TRANSMEM 132
 FT TRANSMEM 144
 FT TRANSMEM 154
 FT TRANSMEM 176
 FT TRANSMEM 196
 FT TRANSMEM 219
 FT TRANSMEM 220
 FT TRANSMEM 245
 FT TRANSMEM 270
 FT TRANSMEM 271
 FT TRANSMEM 297
 FT TRANSMEM 304
 FT TRANSMEM 328
 FT TRANSMEM 391
 FT TRANSMEM 44
 FT TRANSMEM 44
 FT TRANSMEM 48
 FT TRANSMEM 130
 FT TRANSMEM 139
 FT TRANSMEM 339
 FT TRANSMEM 391 AA: 42686 MW: 28C01B27 CRC32;
 SO SEQUENCE

SSRL_HUMAN Length: 391 February 14, 2000 08:02 Type: P Check: 8652

1 MFPNGTASSP SSSPSPSGS CGEGGSRG GAGADGME PGRASNGT
 51 LSEGGSAII ISFITSVCL VGLGNSMT YILYKARK TANNIYILN
 101 AINDELIMS VPELYTSTL RHMPGALC RLVSEVDVN MFSIYGLTV
 151 LSVDRVAVV HPIKARYR PIVAKVNLG VAVLSLVL PIVESTAA
 201 NSDGTACNM LMPEPAQML VGEVLYTFM GELFVGAIC LCYLIITAKM
 251 RYVALRAGM OKRSEKRT LMVMVWVW VECWMPFVV QLVNVEAD
 301 DATVOLSIV LGVANSCLP ILVGFSLDNF KRSPRIICL SWMDNAEEP
 351 VDYATALKS RAYVEDFOP ENLESGVFR NCTCSRIIT L

11AA_SEQUENCE 1.0 STANDARD: PRT: 391 AA.
 ID SSRL_MOUSE
 AC P30873:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SOMATOSTATIN RECEPTOR TYPE 1 (SSIR) (SRIF-2).
 GN SSRI OR SMSTR.
 OS Mus musculus (mouse)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92108031.
 RA YAMADA Y., POST S.R., WANG K., TAGER H.S., BELL G.I., SEINO S.:
 RT "Cloning and functional characterization of a family of human and
 RT mouse somatostatin receptors expressed in brain, gastrointestinal
 RT tract, and kidney."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:251-255(1992).
 CC -1- FUNCTION: RECEPTOR FOR SOMATOSTATIN WITH HIGHER AFFINITY FOR
 CC SOMATOSTATIN-14 THAN -28. THIS RECEPTOR IS COUPLED VIA PERTUSSIS
 CC TOXIN SENSITIVE G PROTEINS TO INHIBITION OF ADENYLATE CYCLASE. IN
 CC ADDITION IT STIMULATES PHOSPHORYLASE PHOSPHATASE AND NA+/H+
 CC EXCHANGER VIA PERTUSSIS TOXIN INSENSITIVE G PROTEINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: JEJUNUM AND STOMACH.

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81831; AAA58255.1; -
 CC PIR: C41795; C41795.
 CC GCRDB: GCR_0273; -
 CC MGI: MGI:98327; SMSTR1.
 CC PROSITE: PS00237; G-PROTEIN_RECEPTOR: 1.
 CC PFM: PFM:PF00001; 7tm_1; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein;
 CC Multigene family; Lipoprotein; Palmitate.
 CC DOMAIN 1
 CC TRANSMEM 57 84
 CC 1 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 2 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 3 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 4 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 5 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 6 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 7 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 8 (POTENTIAL).
 CC CARBOHYD 4 4
 CC CARBOHYD 48 48
 CC CARBOHYD 130 208
 CC DISULFID 339 339
 CC BY SIMILARITY.
 CC PALMITATE (POTENTIAL).
 CC SEQUENCE 391 AA; 42718 MW; BD335205 CRC32;
 SSRL_MOUSE Length: 391 February 14, 2000 08:02 Type: P Check: 8110 ..

1 MPPNGTASSP SSSPSPSGS CGEGACSRG GSGADGME PGRNASQNGT
 51 LSEGGSSAIL ISFIYSVCL VGLGNSVY YVILRYAKK TANNIYLNL
 101 AIADELIMLS VPELVYSTLL RHMPFGALLC RLVSVDVAVN MFTSIYCLIV
 151 LSVDRYAVV HPKARYRR PTVAKVNVNG VWVLSLVYL PIVVESRTAA
 201 NSDGTACNM LMPEPAQRML VGEVLYTFML GFLPEVGAIC LCYVLIANK
 251 RMVALKAGN OQRSEKRT LMVMMVYMF VICMPFYV QLVNVEAEO
 301 DAVSQSLAVI LGYANSCAP ILVGFSLDNF KRSFORILCL SMMDNAEEP
 351 VDYATATLKS RAYSVEDFOP ENLESGGVFR NGTCASRIST L

11AA_SEQUENCE 1.0 STANDARD: PRT: 391 AA.
 ID SSRL_RAT
 AC P28646;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE SOMATOSTATIN RECEPTOR TYPE 1 (SSIR) (SRIF-2).
 GN SSTRI.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WISTAR; TISSUE-BRAIN;
 RX MEDLINE: 92096119.

RA MEYERHOF W., PAUST H.J., SCHOENROCK C., RICHTER D.;
 RT "Cloning of a cDNA encoding a novel putative G-protein-coupled
 RT receptor expressed in specific rat brain regions.";
 RT DNA Cell Biol. 10:689-694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 93016064.
 RA LI X.-J., FORTE M., NORTH R.A., ROSS C.A., SNYDER S.H.;
 RT "Cloning and expression of a rat somatostatin receptor enriched in
 RT brain.";
 RT J. Biol. Chem. 267:21307-21312(1992).
 CC -1- FUNCTION: RECEPTOR FOR SOMATOSTATIN WITH HIGHER AFFINITY FOR
 CC SOMATOSTATIN-14 THAN -28. THIS RECEPTOR IS COUPLED TO
 CC PHOSPHOTRANSFERASE PHOSPHATASE AND NA+/H+ EXCHANGER VIA PERTUSSIS
 CC TOXIN INSENSITIVE G PROTEINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: BRAIN, PITUITARY, ISLET, JEJUNUM, STOMACH,
 CC HEART, SPLEEN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M62314; CAA44193.1; -
 CC EMBL: M97656; -; NOT_ANNOTATED_CDS.
 CC PIR: A39297; A39297.
 CC PIR: A43102; A43102.
 CC GCRDB: GCR_0183; -
 CC GCRDB: GCR_0628; -
 CC PROSITE: PS00237; G-PROTEIN_RECEPTOR: 1.
 CC PFM: PFM:PF00001; 7tm_1; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein;
 CC Multigene family; Lipoprotein; Palmitate.
 CC DOMAIN 1
 CC TRANSMEM 57 84
 CC 1 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 2 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 3 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 4 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 5 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 6 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 7 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 8 (POTENTIAL).
 CC CARBOHYD 4 4
 CC CARBOHYD 48 48
 CC CARBOHYD 130 208
 CC DISULFID 339 339
 CC BY SIMILARITY.
 CC PALMITATE (POTENTIAL).
 CC SEQUENCE 391 AA; 42746 MW; 8AF645D5 CRC32;
 SSRL_RAT Length: 391 February 14, 2000 08:02 Type: P Check: 9240 ..

1 MPPNTASSP TSSPSSPGS CGEGVCSRG GSGADGME PGRNSQNGT
 51 LSEGGSSAIL ISFIYSVCL VGLGNSVY YVILRYAKK TANNIYLNL
 101 AIADELIMLS VPELVYSTLL RHMPFGALLC RLVSVDVAVN MFTSIYCLIV
 151 LSVDRYAVV HPKARYRR PTVAKVNVNG VWVLSLVYL PIVVESRTAA
 201 NSDGTACNM LMPEPAQRML VGEVLYTFML GFLPEVGAIC LCYVLIANK

251 RYVALKAGMO ORKRSERKIT LMVMMVYVME VICMPEFYV QLVNVEAEO
 301 DATVSQUSVI LGVANSKANP ILVYGLSDNF KRSFORILCL SMONNAEKP
 351 VDYATATLKS RAYVEDROP ENLESGVFR NGTCASRIST L

11AA_SEQUENCE 1.0 STANDARD: PRT: 183 AA.
 ID SSRB_CANPA
 AC P23438;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE TRANSLOCON-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)
 DE (SIGNAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA) (GP25H).
 GN SSR2.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-50 AND 83-120.
 RX MEDLINE; 91115924.
 RA GOERLICH D., PREHN S., HARTMANN E., HERZ J., OTTO A., KRAFT R.,
 RA WIEDMANN M., KNESEPEL S., DOBERSTEIN B., RAPPOPORT T.A.;
 RT "The signal sequence receptor has a second subunit and is part of a
 RT translocation complex in the endoplasmic reticulum as probed by a
 RT bifunctional reagents.";
 RL J. Cell Biol. 111:2283-2294(1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-42; 59-106 AND 140-162.
 RX MEDLINE; 92011761.
 RA WADA I., RINDRESS D., CAMERON P.H., CHU W.-J., DOHERTY J.J., II,
 RA LOOYARD D., BELL A.W., DYKARD D., THOMAS D.Y., BERGENSON J.J.M.;
 RT "SR alpha and associated calnexin are major calcium binding proteins
 RT of the endoplasmic reticulum membrane.";
 RL J. Biol. Chem. 266:19599-19610(1991).
 CC -1- FUNCTION: TRAP PROTEINS ARE PART OF A COMPLEX WHOSE FUNCTION IS TO
 CC BIND CA(2+) TO THE ER MEMBRANE AND THEREBY REGULATE THE RETENTION
 CC OF ER RESIDENT PROTEINS.
 CC -1- SUBUNIT: HETEROTETRAMER OF TRAP-ALPHA, TRAP-BETA, TRAP-DELTA AND
 CC TRAP-GAMMA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: X53529; CAA37609.1; -;
 DR EMBL: X53591; CAA37661.1; -;
 DR PIR: A36679; A36679.
 DR PIR: B37273; B37273.
 DR PIR: S15510; S15510.
 KW Glycoprotein; Signal; Endoplasmic reticulum; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 1 183
 FT CHAIN 1 183
 FT DOMAIN 18 149
 FT TRANSSEM 150 169
 FT DOMAIN 170 183
 FT CARBOHYD 88
 FT CARBOHYD 104
 FT SEQUENCE 183 AA; 20100 MW; 977848EB CRC32;
 SSRB_CANPA Length: 183 February 14, 2000 08:02 Type: P Check: 8392

1 MRLASVYLA LFAVSHAEBG ARLLAKSKSL NRYAVEBDL TLQYNIYNG
 51 SSAALDVELS DDSFPEDFG IVSGMLNVKM DRIPASNVS HTVYLRPLKA

101 GYFNFTSATV TYLAOEDGPV VIGFTSAPGO GGLAOREFD RRPSPHFLDM
 151 AAFGVMTLPS IGIPILLWYS SKRKYDTPKS KKN

11AA_SEQUENCE 1.0 STANDARD: PRT: 183 AA.
 ID SSRB_HUMAN
 AC P43308;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE TRANSLOCON-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)
 DE (SIGNAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA).
 GN SSR2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE-COLON CARCINOMA;
 RX MEDLINE; 94114564.
 RA BODESCOT M., BRISON O.;
 RT "Cloning and sequence analysis of the beta subunit of the human
 RT transloccon-associated protein.";
 RL Biochim. Biophys. Acta 1217:101-102(1994).
 CC -1- FUNCTION: TRAP PROTEINS ARE PART OF A COMPLEX WHOSE FUNCTION IS TO
 CC BIND CA(2+) TO THE ER MEMBRANE AND THEREBY REGULATE THE RETENTION
 CC OF ER RESIDENT PROTEINS.
 CC -1- SUBUNIT: HETEROTETRAMER OF TRAP-ALPHA, TRAP-BETA, TRAP-DELTA AND
 CC TRAP-GAMMA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC DR EMBL: X74104; CAA52207.1; -;
 DR MIM: 600867; -;
 KW Glycoprotein; Signal; Endoplasmic reticulum; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 1 183
 FT CHAIN 1 183
 FT DOMAIN 18 149
 FT TRANSSEM 150 169
 FT DOMAIN 170 183
 FT CARBOHYD 88
 FT CARBOHYD 104
 FT SEQUENCE 183 AA; 20135 MW; B2E82B37 CRC32;
 SSRB_HUMAN Length: 183 February 14, 2000 08:02 Type: P Check: 8093

1 MRLSEVYLA LFAVYQAEBG ARLLAKSKSL NRYAVEGDL TLQYNIYNG
 51 SSAALDVELS DDSFPEDFG IVSGMLNVKM DRIPASNVS HTVYLRPLKA
 101 GYFNFTSATV TYLAOEDGPV VIGFTSAPGO GGLAOREFD RRPSPHFLDM
 151 AAFGVMTLPS IGIPILLWYS SKRKYDTPKT KKN

11AA_SEQUENCE 1.0 STANDARD: PRT: 214 AA.
 ID ST14_SOLTU
 AC Q41495;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE STS14 PROTEIN PRECURSOR.
 GN STS14.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CC euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Asteridae; easterids I; Solanales; Solanaceae;
 CC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, DAPIRA: TISSUE-PISTIL;
 RX MEDLINE: 96197407
 RA VAN ELDIK G.J., WIGLEMS G.J., RUTTER R.K., VAN HERPEN M.M.A.,
 RA SCHRUMEN J.A.H., WIGLEMS G.J.,
 RT "Molecular analysis of a pistil-specific gene expressed in the stigma
 and cortex of Solanum tuberosum."
 RL Plant Mol. Biol. 30:171-176(1996).
 CC -1- FUNCTION: MAY PROTECT THE OUTER TISSUES OF THE PISTIL FROM
 CC PATHOGEN ATTACK.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE STIGMA AND STYLAR
 CC CORTEX THROUGHOUT PISTIL DEVELOPMENT. NOT EXPRESSED IN OTHER
 CC ORGANS.
 CC -1- DEVELOPMENTAL STAGE: ACCUMULATES IN THE PISTIL AROUND 120 HOURS
 CC BEFORE ANTHESIS AND INCREASES TOWARDS THE END OF FLOWER
 CC DEVELOPMENT, WITH A MAXIMUM AT ANTHESIS.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X82652; CA57976.1; -
 DR HSP: P04284; ICFE
 DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; FALSE_NEG.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; FALSE_NEG.
 DR PFM: PF00186; SCP; 1.
 KM Repeat: Signal.
 FT CHAIN 1 19 POTENTIAL.
 FT REPEAT 20 214 STS14 PROTEIN.
 FT DOMAIN 13 19 3.5 X 2 AA TANDEM REPEAT OF Y-I.
 FT DOMAIN 59 65 POLY-PRO.
 SQ SEQUENCE 214 AA: 23887 MW; 06BC0717 CRC32;
 SUFL SOLTU Length: 214 February 14, 2000 08:02 Type: P Check: 5285 ..
 1 MEVLSTAMAC LVYIYIYD EKKREKLKVR NKMILTFQ FLITLASSL
 51 THISAQVPP PPPPTSNAT PPSRAQEF L DAHNKARSEV GVGPLTWSPM
 101 LAKETSLVR YORDKONCF ANLSNGKYG NQMAVSGTVV TPRMVDSDW
 151 AKKRYVEN NCTGDDKCG VYQIYWKKS IELGACHTC YEPALTYC
 201 FYNPCNVIG EKP
 11AA SEQUENCE 1.0
 ID SUFL_HAFLIN STANDARD: PRT: 311 AA.
 AC P44847.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SUFL PROTEIN HOMOLOG PRECURSOR.
 GN SUFL OR H10733.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE: 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERAVAGE A.R., BULF C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 MCKENNEY K., SUTTON G., FITZHIGH W., FIELDS C.A., GOCALINE J.D.,

RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTERBACH T.R., HANNA M.C., NGUYEN D.T., SAIDER D.M., BRANDON R.C.,
 RA FINE L.D., FRICHMAN J.L., EHRMAN J.L., GEORGE N.S.M.,
 RA GHEM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.,
 RT "Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd."
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: STRONG, TO E.COLI AND S.TYPIHIMRIUM SUFL.
 CC -1- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U32756; AAC22390.1; -
 DR TIGR: H10733; -
 DR Periplasmic; signal.
 KW SIGNAL 1 28 BY SIMILARITY.
 FT CHAIN 29 311 SUFL PROTEIN HOMOLOG.
 FT DOMAIN 68 164 PLASTOCYANIN-LIKE.
 SQ SEQUENCE 311 AA: 34496 MW; 8065147D CRC32;
 SUFL HAFLIN Length: 311 February 14, 2000 08:02 Type: P Check: 4 ..
 1 MPRLSROL KTAISTALS TVPAPLIAAS REKLVPPLI EVRGRIPLV
 51 TMOETNPBLD GSHNTWGF NGNYLGPTIK IKSSEFALIN YHNNLPOVA
 101 LSIQGLASG ELEGGAARVL KKGESMAPIV PIEOPASCV YRSATLANSA
 151 YQYRGLAGM WLEDEQSLK ANLEPKGYVD DIPLIQDM EFNQGLQLFK
 201 QNCPHEVGNR LVNGIEAPY LDVARGWIRL RLNSIARA YDLRLDNDQ
 251 MLLAODLGF LPAKSVKSL VLSGERAEI LVNMKLITY LSLAEVAVAC
 301 TKNKYVLR R
 11AA SEQUENCE 1.0
 ID SUH3_RAT STANDARD: PRT: 285 AA.
 AC P50235;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ALCOHOL SULFOTRANSFERASE (EC 2.8.2.2) (HYDROXYSTEROID
 SULFOTRANSFERASE) (ST) (ST-60).
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
 RX MEDLINE: 94306585.
 RA WATABE T., OGURA K., SATSUBARA M., OKUDA H., HIRATSUKA A.,
 RT "Molecular cloning and functions of rat liver hydroxysteroid
 sulfotransferases catalysing covalent binding of carcinogenic
 RT polycyclic arylmethanols to DNA."
 RL Chem. Biol. Interact. 92:87-105(1994).
 CC -1- FUNCTION: CATALYSES SULFATION OF HYDROXYSTEROIDS AND XENOBIOTICS
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLSULFATE + AN ALCOHOL -
 CC ADENOSINE 3',5'-BISPHOSPHATE + AN ALKYL SULFATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- INDUCTION: INDUCED BY ESTROGENS AND SUPPRESSED BY ANDROGENS.
 CC EXPRESSION IS UNDER THE INFLUENCE OF PITUITARY GROWTH HORMONE AND
 CC THYROID HORMONE.

CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D14989; BAA03634.1; -
 CC DR PFAM: PF00685; Sulfotransferase; 1.
 CC DR Transferase; Steroid metabolism.
 CC FT BINDING 249 255 PAPS BINDING SITE (POTENTIAL).
 CC SEQUENCE 285 AA; 33531 MW; D4FC185 CRC32;
 SO SUB3_RAT Length: 285 February 14, 2000 08:02 Type: P Check: 800 ..
 1 MMSDYTFEG IPPPAFWFSK EILENSCKKF VKEDDILIT TYPKSGTNL
 51 IETVLIQTK GDPKWIOMSP IMDRSPWIFT GSGYDKLTKM EGPRLMTSHL
 101 PHLFSKSLF SSRKAVIYLI RNPRLVYSA YFFMSKIALE KRPDSIGTY
 151 EWFLEGNVAY GSWFEHIRGW LSNREWDNLF VLYEDMKKD TMSIKKICD
 201 FLGKLEPDE LNLVLYKTSF QVAKENNMSN YSLMEKELIL TGFTFRKGT
 251 TDMKNHFTV AQAEAFDKVF QEKMAGFPPG MFPWE
 11AA_SEQUENCE 1.0
 ID TC21_CHV13 STANDARD; PRT; 278 AA.
 AC P52283;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE TYPE II RESTRICTION ENZYME CVIJI (EC 3.1.21.4) (ENDONUCLEASE CVIJI)
 DE (R.CVIJI).
 OS Chlorella virus IL-3A (CV-IL3A).
 OS Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95331597.
 RA SKOMRON P.M., SWAMINATHAN N., MCMASTER K., GEORGE D., VAN ETEN J.L.,
 RA MEAD D.A.;
 RT "Cloning and applications of the two/three-base restriction
 RT endonuclease R.CviJI from IL-3A virus-infected Chlorella";
 RL Gene 137:37-41(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE: 96292246.
 RA SWAMINATHAN N., MEAD D.A., MCMASTER K., GEORGE D., VAN ETEN J.L.,
 RA SKOMRON P.M.;
 RT "Molecular cloning of the three base restriction endonuclease R.CviJI
 RT from eukaryotic Chlorella virus IL-3A";
 RL Nucleic Acids Res. 24:2463-2469(1996).
 CC -1- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE RGCY
 CC AND CLEAVES AFTER G-2. IN THE PRESENCE OF ATP, THERE IS A
 CC RELAXATION OF ITS SPECIFICITY AND IT CAN CLEAVE RCGN AND YGCY, BUT
 CC NOT YGCR.
 CC -1- COFACTOR: REQUIRES MAGNESIUM.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U09001; AAC55064.1; -
 CC DR REBASE: RB00873; CviJI
 CC DR Hydrolyase; Endonuclease; Nuclease; Restriction system; Magnesium.
 KW

SO SEQUENCE 278 AA; 31633 MW; D23193CE CRC32;
 TC21_CHV13 Length: 278 February 14, 2000 08:02 Type: P Check: 7233 ..
 1 MEEKRIALI EKORLAEEKI ASGRKIKRI SINATHERE FVKYNSAFV
 51 GPATFVVDI KNNKSBEIHN VVREPOLOGS KAKSPAYVD REYNPKRADI
 101 AAVDITKDV AMISHKASEG YQOYLKISG NKFTGKELE EYLSFKRRV
 151 SMAPSKIMP ANKTVSPK SNLIKNOAIF GEDYCKPKR DNVDIGGR
 201 PIIRKRSIL YLFTGFSAI NGHLENFTGK HEPVYVTE RSSGSRITTT
 251 VVNGVYTKNL REFIHPYNEV SSKTORIM
 11AA_SEQUENCE 1.0
 ID TC21_HUMAN STANDARD; PRT; 269 AA.
 AC P36402;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL-SPECIFIC TRANSCRIPTION FACTOR 1 (TCF-1) (T-CELL FACTOR 1)
 DE (TRANSCRIPTION FACTOR-7).
 GN TCF7 OR TCF1
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OS Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91114695.
 RA VAN DE WETERING M.L., OOSTERWEGEL M.A., DOOIJES D., CLEVERS H.;
 RT "Identification and cloning of TCF-1, a T lymphocyte-specific
 RT transcription factor containing a sequence-specific HMG box";
 RL EMBO J. 10:123-132(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92235082.
 RA VAN DE WETERING M.L., OOSTERWEGEL M.A., HOLSTEGE F., DOOIJES D.,
 RA SUJCKERBUIJK R., GEURTS VAN KESSEL A., CLEVERS H.;
 RT "The human T cell transcription factor-1 gene. Structure,
 RT localization, and promoter characterization";
 RL J. Biol. Chem. 267:8530-8536(1992).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR INVOLVED IN T-CELL LYMPHOCYTE
 CC DIFFERENTIATION. IT BINDS TO THE T-LYMPHOCYTE-SPECIFIC ENHANCER
 CC ELEMENT (5'-WCAAG-'3') FOUND IN THE PROMOTER OF THE CD3-EPSILON
 CC GENE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: THREE DIFFERENT FORMS THAT DIFFER IN THE
 CC C-TERMINAL SECTION ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: T-CELL SPECIFIC.
 CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59869; CAA42526.1; -
 CC DR EMBL: X59870; CAA42527.1; -
 CC DR EMBL: X59871; CAA42528.1; -
 CC DR EMBL: X63901; CAB56795.1; -
 CC PIR: S16645; S16645.
 CC PIR: A38095; A38095.
 CC TRANSFAC: T00999; -
 CC DR TRANSFAC: T01001; -
 CC DR MIM: 189908; -
 CC DR PFAM: PF00505; HMG_Box; 1.
 CC DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
 CC Alternative splicing.
 KW

FT DNA_BIND 154 222 HMG_BOX.
 FT VARSPPLIC 244 269 ETNMPRELKDGNGEISLSSSSSSPA -> GGRNAFTYPT
 FT VARSPPLIC 244 269 EKAAPAPFLPMYVL (IN ISOFORM B).
 FT VARSPPLIC 244 269 ETNMPRELKDGNGEISLSSSSSSPA -> DGGSPKCRAR
 FT SEQUENCE 269 AA; 30264 MW; 7BB387D9 CRC32;
 TCF1_HUMAN Length: 269 February 14, 2000 08:02 Type: P Check: 5007

1 MKKEVYSAF NLMMHYPPPS GAGQHPPQP PLHKANQPH GYQSLSYEH
 51 FNSPHPTAP ADISQKQVHR PLQTPDLSEF YSLTSGSMGO LPHTVWFTH
 101 PSLMIGSGVP GHRAIPHPA IYPPSGKQL QPFDRLNLTQ AESKAKKEAK
 151 KTIKIKPLNA PMLYMKEMRA KYIAECTLKE SAALNQLIGR RWHALSREEO
 201 AAYTELARKE ROLHMQLYPC WSAADNYGKK KRSRERXHOE STETNMPRE
 251 LKDGNGQSL SMSSSSSPA

11AA_SEQUENCE 1.0
 ID TEST_HUMAN STANDARD; PRT; 167 AA.
 AC 015273;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE TELETHONIN (TITIN CAP PROTEIN).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.

[1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE; 98010471.
 RA VALLE G., FALKNER G., DE ANTONI A., PACCHIONI B., PALLAVICINI A.,
 RA PANDOLFO D., TISO N., TOPPO S., TREVISAN S., LANFRANCHI G.,
 RT "telethonin, a novel sarcomeric protein of heart and skeletal
 RT muscle.";
 RL FEBS Lett. 415:163-168(1997).

[2]
 RP SEQUENCE FROM N.A.
 RA PALLAVICINI A., VALLE G., LANFRANCHI G.;
 RT "Human telethonin genomic sequence.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

[3]
 RP SEQUENCE FROM N.A.
 RA MOES A., GAUTEL M.;
 RT "Structure of the human telethonin gene.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

[4]
 RP INTERACTION WITH TITIN.

RA MEDLINE; 98307394.
 RA MOES A., VAN DER VEN P.F., YOUNG P., FURST D.O., GAUTEL M.;
 RT "Two immunoglobulin-like domains of the z-disc portion of titin
 RT interact in a conformation-dependent way with telethonin.";
 RL FEBS Lett. 428:111-114(1998).
 CC -1- FUNCTION: MUSCLE ASSEMBLY REGULATING FACTOR.
 CC -1- SUBUNIT: INTERACTS WITH TITIN.
 CC -1- TISSUE SPECIFICITY: HEART AND SKELETAL MUSCLE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AJ000491; CAA04129.1;
 DR EMBL; AJ010063; CAA08987.1;
 DR EMBL; AJ011098; CAA09479.1;

SO SEQUENCE 167 AA; 19052 MW; EBCDB76C CRC32;

TELT_HUMAN Length: 167 February 14, 2000 08:02 Type: P Check: 3375

1 MATSELSCEV SEENGERREA FWAEMKDLTL STREPEGCSL HEEDTORHET
 51 YHQOGCCOVL VQSPWLMNR MGILGRGLOE YOLPYQVLP LPITPAKMG
 101 AKTEREDTP IQLRELLALE TALGGCCVDR QDAVEITKOL PVPVYSKPG
 151 ALRRSLRSM SOEAQRG

11AA_SEQUENCE 1.0
 ID TEST_MOUSE STANDARD; PRT; 167 AA.
 AC 070348;

DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE TELETHONIN (TITIN CAP PROTEIN).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

[1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-DIAPHRAGM;
 RA IVOLELLA C., FORRENTIN E., VALLE G., LANFRANCHI G.;
 RT "Skeletal muscle transcripts characterization in Homo sapiens and Mus
 RT musculus.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

[2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HEART;
 RA KOEMERER B.;
 RT "The titin cap protein - a novel protein essential for sarcomere
 RT formation.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[3]
 RP -1- FUNCTION: MUSCLE ASSEMBLY REGULATING FACTOR.
 CC -1- SUBUNIT: INTERACTS WITH TITIN.
 CC -1- SUBUNIT: INTERACTS WITH TITIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AJ223854; CAA11585.1;
 DR EMBL; Y15845; CAB38077.1;
 SO SEQUENCE 167 AA; 19078 MW; 06F264F1 CRC32;

TELT_MOUSE Length: 167 February 14, 2000 08:02 Type: P Check: 5613

1 MATSELSCEV SEENGERREA FWAEMKDLTL STREPEGCSL HEEDTORHET
 51 YHQOGCCOVL VQSPWLMNR MGILGRGLOE YOLPYQVLP LPITPAKMG
 101 ASKEEREETP IQLRELLALE TALGGCCVDR QDAVEITKOL PVPVYSKPG
 151 PLRRLSRSM SOEAQRG

11AA_SEQUENCE 1.0
 ID TRAM_AGR6 STANDARD; PRT; 102 AA.
 AC 057471;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL REPRESSOR TRAM.

OS Agrobacterium tumefaciens.
 OC Plasmid pTiA6NC.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae: Agrobacterium.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 95173115.
 RA FLOU C., BURBEA M., WINANS S.C.;
 RT "Activity of the Agrobacterium T1 plasmid conjugal transfer regulator
 RT TR is inhibited by the product of the *traM* gene";
 RT J. Bacteriol. 177:1367-1373(1995).
 CC -1- FUNCTION: NEGATIVELY REGULATES CONJUGATION AND *TRA* GENES
 CC EITHER BIND OR MODIFY *TRAR* OR *RAI* MAKING THEM UNAVAILABLE
 CC ALTERNATIVELY, *TRAM* MAY BIND *TRA* PROMOTERS PREVENTING *TRAR*
 CC ACTIVATION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U43674; AAC28120.1; -
 DR EMBL: U16786; AAA64838.1; -
 KW Transcription regulation; Repressor; Plasmid.
 SQ SEQUENCE 102 AA; 11401 MW; AAC3EDD6 CRC32;
 TRAM_AGR6 Length: 102 February 14, 2000 08:02 Type: P Check: 4597 ..
 1 MELEBANATK KVELRPLIG TRGLPTDLE TITIDARTR RLVEKADL
 51 FOALRETYRT GQACGPGPHI RYLEASIEH AOMSAINTLY SILGFIPKV
 101 VN
 !!AA_SEQUENCE 1.0
 ID TRM_SYNY3 STANDARD; PRT; 231 AA.
 AC P72828;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31) (MIG-
 DE METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE).
 GN TRM OR SL1198.
 OS Synchocystis sp. (strain PCC 6803).
 OC Bacteria: Cyanobacteria; Chroococcales; Synchocystis.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97061201.
 RA KANeko T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
 RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
 RA YAMADA M., YASUDA M., TABATA S.,
 RT Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.;
 RL DNA Res. 3:109-136(1996).
 CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-
 CC L-HOMOCYSTEINE + TRNA CONTAINING N(1)-METHYLGUANINE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRM FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D90901; BAA16843.1; -
 KW Transferase; Methyltransferase; tRNA processing.
 SQ SEQUENCE 231 AA; 25949 MW; B0CF12H1 CRC32;
 TRM_SYNY3 Length: 231 February 14, 2000 08:02 Type: P Check: 6210 ..
 1 MGFVLLTFP DEFTSPLOSG LLGKALEKAI ASYNLINPRD FTTKHRRVD
 51 DEFGGAGVM VIKPEIFAA VESLPVLSK EYLMTPOG PMDOLFREL
 101 TNYQVLIC GHYGVDERV COLVREYSL GDEVLCGEI PALTINGVI
 151 RLLEGTGKE ASLIASFST DLDYPHTR PPFVRLAVP PYLLSGNHQA
 201 IAWRLRQOE ERQQRPRDL WQKWDNRPS P
 !!AA_SEQUENCE 1.0
 ID TRM_SCHRO STANDARD; PRT; 415 AA.
 AC O13947;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
 DE (EC 2.1.1.61).
 GN SPAC23H4.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA BROWN D.; CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-L-
 CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
 CC THIOURIDYLATE.
 CC -1- SIMILARITY: BELONGS TO THE TRM FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z98977; CAB11659.1; -
 DR Transferase; Methyltransferase; tRNA processing.
 KW SEQUENCE 415 AA; 47626 MW; 36B7C485 CRC32;
 TRM_SCHRO Length: 415 February 14, 2000 08:02 Type: P Check: 7997 ..
 1 MVSFLQKQ IIGSKAFQF HSTRLOWPKS QDKVFVMSG GVDSSFSAYL
 51 LKQSYNVEG VPKRNWDED SAPSCGPAER DMAIVQVCK KINISRRRN
 101 FKEKWNLVF EPSLIDYENG LTPNDVSCN RQVFGALFD ALKKHENVV
 151 KQDWMLASGH YAKSVNIET NESHCIPFD KRKQITFLC TIRKALEKT
 201 IEPHLNWKIE NVKKQASSAG FEELAKQDS QGLCFVSPNV GKKRFKFLR
 251 YLNFSDRIK VIAGKNVGE FSGNGHISL TVSGRCGLSL POAOSFYGR
 301 WYVWKDKIN NALYICRGT NELMSKCIY LKDKKWCGR LONLEKSAIS
 351 CEVVRHOOP LOPAKYTRN PESVRIHFD KQRAVTPGOV IAVYNDVCL
 401 GGGVDTVEP EKDFD
 !!AA_SEQUENCE 1.0
 ID TRM_YEAST STANDARD; PRT; 417 AA.

AC 012093;
AT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
(EC 2.1.1.61).
GN YD1033C OR D2761.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN: S288C;
RA PAULIN L., SAREN A.M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA - S-ADENOSYL-L-
CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
CC THIOURIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE TRMU FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z71781; CAA96456.1; -
CC EMBL: Z74081; CAA98591.1; -
CC DR Transferase: Methyltransferase; trna processing.
CC KW SEQUENCE 417 AA; 47049 MW; 4690E911 CRC32;
SQ
TRMU_YEAST Length: 417 February 14, 2000 08:02 Type: P Check: 8415 ..

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20475; AAA33280.1; -
CC DR HSSP: P19992; 2HSD.
CC DR PROSITE: PS00061; ADH_SHORT; 1.
CC DR PFAM: PF00106; adh_short; 1.
CC DR PFAM: PF00678; adh_short_C2; 1.
CC KW Oxidoreductase; NADP.
CC FT ACT_SITE 21 45 NADP (BY SIMILARITY).
CC FT ACT_SITE 167 167 BY SIMILARITY.
CC SQ SEQUENCE 268 AA; 28588 MW; 0BB5A16C CRC32;
TRNH_DATST Length: 268 February 14, 2000 08:02 Type: P Check: 5293 ..

1 MAGREIIGGD RRMSLRGMTA LVTGTRGIG VAIVEELANF GAELYTCSSRS
51 QNDLDECLER WRKRGFKVSG PCDVDSISG RQTLMEVTS SPNGKINILI
101 NNAGTIIPKE AINFTAEYS IINGTNEAS YVLCOLAPL LKASGNASIV
151 FNSNAGVIA VELSITVYAS KGAINOVTKS LACEBAKSI RVANVAPWII
201 NTFIIEAQQ VPSQKNIES LIGRAPAKRA GEPSEVSIV TYLCPLASTY
251 ITGQIICVDG GYTVNGFI

11AA_SEQUENCE 1.0 STANDARD; PRT; 312 AA.
ID TRPE-CRYNE
AC P27710;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
GN N-(5'-PHOSPHORIBOSYL)ANTHRANILATE ISOMERASE (EC 5.3.1.24) (PRAI).
OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellales;
OC Tremellaceae; Filobasidiella.
CC [1]
CC SEQUENCE FROM N.A.
CC RX MEDLINE: 93083987.
CC RA PERFECT J.R., RUDE T.H., PENNING L.M., JOHNSTON S.A.;
CC RT "Cloning the cryptococcus neoformans TRP1 gene by complementation in
CC RT Saccharomyces cerevisiae."
CC RL Gene 123:213-217(1993).
CC -1- CATALYTIC ACTIVITY: N-(5'-PHOSPHO-BETA-D-RIBOSYL)-ANTHRANILATE -
CC 1-(2-CARBOXYPHENYLAMINO)-1-DEOXY-D-RIBULOSE 5-PHOSPHATE.
CC -1- PATHWAY: THIRD STEP IN TRIPHOHAN BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M74901; AAA51445.1; -
CC DR PIR: JN0451; JN0451.
CC DR HSSP: O56320; INST.
CC DR PFAM: PF00218; IGPS; 1.
CC DR PFAM: PF00697; PRAI; 1.
CC KW Isomerase; tryptophan biosynthesis.
CC SQ SEQUENCE 312 AA; 33070 MW; CSA4F8A8 CRC32;

Page 127

```

REF_CRYNE Length: 312 February 14, 2000 08:02 Type: P Check: 2342 ..

1 MSTSVMNNA LNGRDVLCA LSGISSEHDV EKYKEGVKG VLVEALMRA
51 SDTKALRSL IGLPLEYVS KPRLYVICG INSTDAKLA INAGADLLGV
101 ILVPGTRRCI STSTAREISA LVOSARSOSS SKPLPSLSS PMFQSQALL
151 SSRKRRLNG VFONOSLSDI LSANDEIGLD LVQHDEPO AMAKFIPIPV
201 VKYFRVSPBS IVRGGEIRP GLNALILDA GGASGGGGG KAFWEHAKR
251 LIOSGEVGE GHVLPYLILA GGLTPENVG ALNRLYKANG VMMSAVGSKE
301 REGRRRRRLR RS

IIA_SEQUENCE 1.0 STANDARD; PRT: 371 AA.
ID TTF1_CANFA
AC P43698;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THYROID TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
DE (HOMEBOX PROTEIN NKX-2.1).
DE TTF1 OR TTF1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96034516.
RA VAN RENTERGHEM P.H.G., DREMER S., VASSAR G., CHRISTOPHE J.;
RT "Study of TTF-1 gene expression in dog thyrocytes in primary
RT culture."
RT Mol Cell. Endocrinol. 112:83-93(1995).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
CC THYROPEROXIDASE AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
CC ROLE IN LONG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
CC -1- SUBCELLULAR LOCATION: NUCLEOL.
CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND CNS.
CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77910; CAA54868.1; -
DR HSSP: P23441; 1FTT.
DR TRANSFAC: T02098; -
DR PROSITE: PS00027; HOMEBOX.1; 1.
DR PROSITE: PS50071; HOMEBOX.2; 1.
DR PFAM: PF00046; homeobox; 1.
KW Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein.
FT DNA BIND 161 220 HOMEBOX.
FT DOMAIN 234 243 POLY-GLY.
FT DOMAIN 246 253 POLY-GLN.
FT DOMAIN 294 303 POLY-NLA.
SQ SEQUENCE 371 AA: 38539 MW: 56C6C42D CRC32:
TTF1_CANFA Length: 371 February 14, 2000 08:02 Type: P Check: 4988 ..

1 MSMSPKHTTP EYSVDLSLP DESYKKVME GGLIGAPLAA YKQGQAAPA
51 AAMQGHVGH HGAVTAAYHA TAAGVPQLSH SAVGGYGNNG LGNMSELPPY
101 QDTWNSASG PGWYGANPD REPATSRMG PASGMNSGM GGLIGSLGDV

151 KNMAPLPASP RRKRRLFQSO AQVELERR KOOKYLSAPE RELHSMITHL
201 TPYQVKWFQ NHRYMKRQA KDVAAOOQLO QDSGGGGGG GAGCPQQOQA
251 QQSPRRVAV PVLVKGKPC QAGAPOPAG SIQGHAOOA QOQAQAOAA
301 AAIVSGSG PELGANPHQ PGSAGSPDL AHHAAPAL OGQVSSLHL
351 NSSGSDYGTN SCSTLYGRT W

IIA_SEQUENCE 1.0 STANDARD; PRT: 371 AA.
ID TTF1_HUMAN
AC P43659; O14955; O14954;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THYROID TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
DE (HOMEBOX PROTEIN NKX-2.1).
DE TTF1 OR TTF1 OR NKX2A.
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95226464.
RA SALARDI A., TASSI V., DE FILIPPE V., CIVITAREALE D.;
RT "Cloning and sequence analysis of human thyroid transcription factor
RT 1."
RT Blochim. Biophys. Acta 1261:307-310(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA IKEDA K., CLARK J.C., SHAW-WHITE J.R., STAHLMAN M.T., BOUTELL C.J.;
RA WHITSITT J.A.;
RT "Gene structure and expression of human thyroid transcription
RT factor-1 in respiratory epithelial cells."
RT J. Biol. Chem. 270:8108-8114(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA OUCHI H., KIMURA S.;
RT Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA HAMADA H., LIU H., DELENOS R., MINOO P.;
RT Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA HAMADA H., LIU H., JONES C., DELENOS R., MINOO P.;
RT Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA ENDO T., OHNO M., NAKAZATO M.;
RT Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
CC ROLE IN LONG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
CC -1- SUBCELLULAR LOCATION: NUCLEOL.
CC -1- TISSUE SPECIFICITY: THYROID AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X82850; CAAS8053.1; -
 CC EMBL: U19816; AAC50125.1; -
 DR EMBL: U19756; AAA86099.1; -
 DR EMBL: U43203; AAA89066.1; ALT_INIT.
 DR EMBL: U33749; AAB52381.1; -
 DR EMBL: D50740; BAA23529.1; -
 DR EMBL: D50739; BAA23527.1; -
 DR HSSP: P23441; 1FTT.
 DR TRANSFAC: T00857; -
 DR MIM: 600635; -
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFM: PF00046; homeobox; 1.
 KW Transcription regulation; Activator; Homeobox; DNA-binding;
 FT DNA_BIND 161 220 HOMEBOX.
 FT DOMAIN 234 243 POLY-GLY.
 FT DOMAIN 246 253 POLY-GLN.
 FT VARSPLIC 112 125 GMYGANDPREFPAT -> V (IN ISOFORM 2).
 FT CONFLICT 49 49 P -> H (IN REF. 6).
 FT CONFLICT 61 61 H -> P (IN REF. 6).
 FT CONFLICT 158 158 S -> T (IN REF. 6).
 FT CONFLICT 161 161 R -> G (IN REF. 6).
 FT CONFLICT 226 227 QQ -> HE (IN REF. 5).
 SQ SEQUENCE 371 AA; 38596 MW; FEA215E7 CRC32;

TF1_HUMAN Length: 371 February 14, 2000 08:02 Type: P Check: 5439

1 MSMSPKHTTP FVSVDLSPL EESYKKYGM GGLGAPLAA YRQGAAPPT
 51 AAMQOHAVGH HGAVTAAYHM TAAGVPLSH SAVGTCNGN LGNMSELPPY
 101 QDTMRNSASG PGWYGANDP RFPALSRFG PASGNNMSG GGLSLGDVS
 151 KMAVLPSP RRRRYLFSQ AQYELERRF KQKYLSP REHLASMTL
 201 TPQYKIMFO NHRYSKMRQA KDKAAQOQL QDSGGGGGG GTCPCPOQA
 251 QOQSPRRVAV PVLYKDKRP QAGAPAPGA SLQGHAAQA QHQQAQAQA
 301 AAAISVSGG AGLGAPGHQ PGSAGSPDL AHHAASPAL QGVVSLSHL
 351 NSSGSDYGT MSCSTLYGRT W

11AA_SEQUENCE 1.0 STANDARD; PRT; 372 AA.
 ID TF1_MOUSE
 AC P50220.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THYROID-TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
 DE (HOMEBOX PROTEIN NKX-2.1).
 GN TTF1 OR TTF1 OR NKX-2.1 OR NKX-2.1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV; TISSUE-LIVER;
 RX MEDLINE: 95226463
 RA OUCHT H., PAN Y.T., KIMURA S.;
 RT "The complete nucleotide sequence of the mouse thyroid-specific
 enhancer-binding protein (T/EBP) gene: extensive identity of the
 deduced amino acid sequence with the human protein.";
 RL Blochim. Biophys. Acta 1261:304-306(1995).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
 CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYOGLOBULIN,
 CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
 CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A

ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U19755; AAA86100.1; -
 CC HSSP: P23441; 1FTT.
 DR TRANSFAC: T00859; -
 DR MGD; MGI:108067; TTF1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFM: PF00046; homeobox; 1.
 KW Transcription regulation; Activator; Homeobox; DNA-binding;
 FT DNA_BIND 161 220 HOMEBOX.
 FT DOMAIN 234 244 POLY-GLY.
 FT DOMAIN 247 254 POLY-GLN.
 FT DOMAIN 295 304 POLY-ALA.
 SQ SEQUENCE 372 AA; 38570 MW; E047C32A CRC32;

TF1_MOUSE Length: 372 February 14, 2000 08:02 Type: P Check: 7622

1 MSMSPKHTTP FVSVDLSPL EESYKKYGM GGLGAPLAA YRQGAAPPT
 51 AAMQOHAVGH HGAVTAAYHM TAAGVPLSH SAVGTCNGN LGNMSELPPY
 101 QDTMRNSASG PGWYGANDP RFPALSRFG PASGNNMSG GGLSLGDVS
 151 KMAVLPSP RRRRYLFSQ AQYELERRF KQKYLSP REHLASMTL
 201 TPQYKIMFO NHRYSKMRQA KDKAAQOQL QDSGGGGGG GAGCPOQOQ
 251 AQQQSPRRVAV PVLYKDKRP QAGAPAPGA ASLSHAQQA AQQQAQAQA
 301 AAAISVSGG GAGLGAPGHQ QPSAGSPDL LAHHAASPAQ LQGVVSLSH
 351 LNSSGSDYGA MSCSTLYGRT W

11AA_SEQUENCE 1.0 STANDARD; PRT; 372 AA.
 ID TTF1_RAT
 AC P23441; O08630; 070121;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THYROID-TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
 DE (HOMEBOX PROTEIN NKX-2.1).
 GN TTF1 OR TTF1 OR NKX-2.1 OR NKX-2.1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 91006063.
 RA GAZZINI S., PRICE M., DE FELICE M., DAMANTE G., MATTEI M.-G.,
 RA DI LAURO R.;
 RT "Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays
 RT a novel DNA binding specificity.";
 RL EMBO J. 9:3631-3639(1990).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 95071477.
 RA ENDO T., OHTA K., SAITO T., HARAGUCHI K., NAKAZATO M., KOGAI T.,
 RA ONAYA T.;
 RT "Structure of the rat thyroid transcription factor-1 (TTF-1) gene.";

RL Biochem. Biophys. Res. Commun. 204:1358-1363(1994).

RN [3]

RP STRUCTURE BY NMR OF 161-227.

RX MEDLINE: 94109576.

RA VIGILINO P., FOCOLARI F., FORMISANO S., BORTOLOTTI N., DAMANTE G.,

RA DI LAURO R., ESPOSITO G.,

RT "Structural study of rat thyroid transcription factor 1 homeodomain (TF1-HD) by nuclear magnetic resonance."

RT FEBS Lett. 336:397-402(1993).

RN [4]

RP STRUCTURE BY NMR OF 161-227.

RX MEDLINE: 97054597.

RA ESPOSITO G., FOCOLARI F., DAMANTE G., FORMISANO S., TELL G.,

RA DI LAURO R., VIGILINO P.,

RT "Analysis of the evolution structure of the homeodomain of rat thyroid transcription factor 1 by 1H-NMR spectroscopy and restrained molecular mechanics."

RT Eur. J. Biochem. 241:101-113(1996).

CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN, THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND CNS. EXPRESSED IN RESTRICTED REGIONS OF THE DEVELOPING BRAIN WITHIN THE Diencephalon, IN PARTS OF THE HYPOTHALAMUS AND NEUROHYPOPHYSIS, AND IN THE TELECEPHALON.

CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X53858; CA37851.1; -

DR EMBL: D38035; BAA07231.1; ALT-SEQ.

DR PIR: S12002; S12002.

DR PDB: 1FTT; 28JAN-96.

DR TRANSLAC: T00835; -

DR PROSITE: PS00027; HOMEBOX_1: 1.

DR PROSITE: PS00071; HOMEBOX_2: 1.

DR PIR: P00046; homeobox; 1.

CC Transcription regulation; Activator; Homeobox; DNA-binding;

CC Nuclear protein; 3D-structure.

FT DNA_BIND 161 220 HOMEBOX.

FT DOMAIN 234 244 POLY-GLY.

FT DOMAIN 247 254 POLY-GLN.

FT DOMAIN 295 304 POLY-ALA.

SO SEQUENCE 372 AA; 38554 MW; AE0F1572 CRC32;

TXF1_DENAN Length: 372 February 14, 2000 08:02 Type: P Check: 6608 ..

1 MSMPKHTTP FSVSDILSPL EESYKKGME GGLGAPLAA YKOGAAPP

51 AAMQOHAVGH HGAVTAAIYHM TAAGVPLSH SAVGCGYNGN LGMSLPPY

101 QDTRNSNSG PGWYGANPD RPAPSRFNG PASGMNSGM GGLGSLGDS

151 KNAAPLPAP RRRRVLFSG AOVYELERF KQKYLISAP REHLASMTL

201 TPTQKIVTFQ NRRYKRRQA KDKAAQOOLQ ODSGGGGGG GAGACPQQOQ

251 AQQGSPRVA VPVLKDGKP COAGAPAPGA ASLQGHAAQO AQQQAQAAQ

301 AAAAISVSG GAGLGAHPGH QGSGASQSD LAHHAASPA LOGGVSSLSH

351 LNSGSDYGA MSCSTLLYGR TW

!!AA_SEQUENCE 1.0 STANDARD: PRT: 81 AA.

ID TXF8_DENAN

AC P01404;

DT 21-JUL-1986 (rel. 01. Created)

DT 01-FEB-1994 (rel. 28. Last sequence update)

DT 01-FEB-1994 (rel. 28. Last annotation update)

DE TOXIN F-VIII PRECURSOR (TOXIN TAZ) (TOXIN DAF8).

OS Dendroaspis angusticeps (Eastern green mamba).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae; Elapidae; Elapinae; Dendroaspis.

CC [1]

CC SEQUENCE FROM N.A.

CC TISSUE-VEINOM GLAND;

CC DUCANCEL F.;

CC Submitted (JUN-1990) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 22-81.

RX MEDLINE: 74070661.

RA VILJOEN C.C., BOTES D.P.;

RT "Snake venom toxins. The purification and amino acid sequence of toxin Taz from Dendroaspis angusticeps venom."

RT J. Biol. Chem. 249:3863-372(1974).

CC -1- FUNCTION: ITS PHARMACOLOGICAL ACTION IS UNKNOWN. IT PROBABLY ACTS SYNERGISTICALLY WITH OTHER VENOM COMPONENTS AS WHOLE VENOM IS VERY TOXIC.

CC -1- MISCELLANEOUS: LD(50) IS >250 MG/KG BY SUBCUTANEOUS INJECTION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X53409; CA37485.1; -

DR PIR: A01675; TSEPZA.

DR PIR: S21299; S21299.

DR HSSP: P01416; INTX.

DR PROSITE: PS00272; SNAKE_TOXIN; 1.

DR PIR: P00087; toxin; 1.

CC Venom; Toxin; Signal

FT CHAIN 1 21

FT CHAIN 22 81

FT DISULFID 24 43 TOXIN F-VIII.

FT DISULFID 38 60 BY SIMILARITY.

FT DISULFID 62 73 BY SIMILARITY.

FT DISULFID 74 79 BY SIMILARITY.

FT DISULFID 54 54 I -> V (IN REF. 2).

FT CONFLICT 54 54

SO SEQUENCE 81 AA; 8849 MW; E0BFC94E CRC32;

TXF8_DENAN Length: 81 February 14, 2000 08:02 Type: P Check: 8703 ..

1 MKTLLTLV VTIYCLDAS TMIYSHKTP OPSATITCEE KTCYKRSYRK

51 LPAIVAGRG GCPKREMLVA IHCRRSDKN E

!!AA_SEQUENCE 1.0 STANDARD: PRT: 153 AA.

ID UBC7_DROME

AC P52487;

DT 01-OCT-1996 (rel. 34. Created)

DT 01-OCT-1996 (rel. 34. Last sequence update)

DT 01-NOV-1997 (rel. 35. Last annotation update)

DE UBIQUITIN-CONJUGATING ENZYME P2-18 KD (EC 6.3.2.19) (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

CC [1]

CC SEQUENCE FROM N.A.

CC ROBIN C., MEDVECKY K.M., RUSSELL R.J., OAKESHOTT J.G.;

RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -
 CC AMP + PYROPHOSPHATE + PROTEIN-N-UBIQUITYLATION.
 CC - PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
 CC - MISCELLANEOUS: A Cysteine residue is required for
 CC UBIQUITIN-THIOLESTER FORMATION (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 CC STRONGEST: TO HUMAN UBC7.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U51051; AAB01150.1; -
 DR HSSP: P15731; 20CE.
 DR ELYBASE: EFGN0017456; UBC84D.
 DR PROSITE: PS00183; UBIQUITIN_CONJUGAT; 1.
 DR PFAM: PF00179; UQ.con; 1.
 DR Ubiquitin conjugation: Llgase.
 KW BINDING: 86
 FT SEQUENCE 153 AA: 17709 MW: 41698C20 CRC32;
 SQ
 UBC7_DROME Length: 153 February 14, 2000 08:02 Type: P Check: 1891 ..
 1 MATRRRLTTE LSDLYEAKMS TLKNISSSE SLMTATGLV PERAPYKGA
 51 XRIENFPO YPEMPKILF KTYIHPND EKGEVCLPIL STDNWKETTR
 101 TEQVIALVA IVHNEPEHP LRSDLAEFV REHKFKMKA EEFTKNAEK
 151 RPE
 !!AA_SEQUENCE 1.0
 ID UMUD_ECOLI STANDARD: PRT; 139 AA.
 AC P04153;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 15-JUL-1999 (Rel. 03, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE UMUD PROTEIN (EC 3.4.21.-) [CONTAINS: UMUD' PROTEIN].
 GN UMUD.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85242679.
 RA KITAGAWA Y., AKASOHI E., SHINAGAWA H., HORII T., OGAWA H., KATO T.;
 RT Structural analysis of the umu operon required for inducible
 RT mutagenesis in Escherichia coli.;
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4336-4340(1985).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85242678.
 RA PERRY K.L., ELLIOTT S.J., MITCHELL B.B., MARSH L., WALKER G.C.;
 RT "umuC and umuAB operons whose products are required for UV light-
 RT and chemical-induced mutagenesis: Umud, MucA, and LexA proteins share
 RT homology.";
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4331-4335(1985).
 RL [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GODDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12."

RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 97061202.
 RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
 RA IKEMOTO K., INABA T., ITOH T., KAJIHARA M., KAWAI K., KASHIMOTO K.,
 RA KITURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
 RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SATO N.,
 RA SANEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
 RA YANO M., HORIOUCHI T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12,728,0 min region on the linkage map.";
 RT DNA Res. 3:137-155(1996).
 RN [5]
 RP MUTAGENESIS.
 RX MEDLINE: 92318898.
 RA KOCH W.H., ENNIS D.G., LEVINE A.S., WOODGATE R.;
 RT "Escherichia coli umud mutants: DNA sequence alterations and umud
 RT cleavage.";
 RT Mol. Gen. Genet. 233:443-448(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF UMUD'.
 RX MEDLINE: 96202734.
 RA PEAT T.S., FRANK E., MCDONALD J.P., LEVINE A.S., WOODGATE R.,
 RA HENDRICKSON W.A.;
 RT "Structure of the umud' protein and its regulation in response to DNA
 RT damage.";
 RT Nature 380:727-730(1996).
 CC - FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION. ESSENTIAL FOR
 CC INDUCED (OR SOS) MUTAGENESIS. MAY MODIFY THE DNA REPLICATION
 CC MACHINERY TO ALLOW BYPASS SYNTHESIS ACROSS A DAMAGED TEMPLATE.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S24 ALSO KNOWN AS THE
 CC UMUD/LEXA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M10107; AAA24728.1; -
 DR EMBL: M13387; AAA98073.1; -
 DR EMBL: AE000216; AAC74267.1; -
 DR EMBL: D90752; BAA36030.1; -
 DR EMBL: D90753; BAA36038.1; -
 DR PIR: A03551; ZWECU.
 DR PIR: A23157; A23157.
 DR PDB: 1UMU; 01-ANG-96.
 DR PDB: 1AY9; 28-JAN-98.
 DR ECO2DBASE: A015.1; 6TH EDITION.
 DR ECOGENE: EG11057; UMUD.
 DR PRAW: PFW0717; Peptidase_S24.1
 KM SOS mutagenesis; DNA repair; Autocatalytic cleavage; Hydrolase;
 KM Serine protease; 3D-structure.
 FT PROPEP 1
 FT CHAIN 25 139
 FT SITE 24 25
 FT ACT_SITE 60 60
 FT ACT_SITE 97 97
 FT MUTAGEN 27 27
 FT MUTAGEN 65 65
 FT FT 92 92
 SQ SEQUENCE 139 AA: 15063 MW: D0175DFA CRC32;
 UUD_ECOLI Length: 139 February 14, 2000 08:02 Type: P Check: 8403 ..
 1 MFIIPALR EITPFLFS LVQCFPSA ADVVEGHID NLLIQHSA
 51 TYFVASDS MIDGISDGD LLYVSALTA SHQDIYAV DSEFTVKLO

101 LRPYVOLIPM NSAYSPTIS SEDTLDVGV VHYVKAMR

11AA_SEQUENCE 1.0 STANDARD; PRT: 303 AA.

ID US02_HSVK

AC 032517;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE US1 PROTEIN.

GN US1.

OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicelloviruses.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 93079867.

RA BREEDEN C.A., YALAMANCHILI R.R., COLLE C.F. III, O'CALLAGHAN D.J.;

RT "Identification and transcriptional mapping of genes encoded at the

RT IR/US junction of equine herpesvirus type 1.";

RL Virology 191:649-660(1992).

RN [2]

RP SEQUENCE OF 1-209 FROM N.A.

RX MEDLINE: 92263758.

RA COLLE C.F. III, FLOWERS C.C., O'CALLAGHAN D.J.;

RT "Open reading frames encoding a protein kinase, homolog of

RT glycoprotein gX or pseudorabies virus, and a novel glycoprotein map

RT within the unique short segment of equine herpesvirus type 1.";

RL Virology 188:345-357(1992).

CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES US2 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M87497; AAA46069.1; -

DR PIR: A44215; A44215. E -> G (IN REF. 2).

FT CONFLICT 31 31

FT MOD.RES 164 164

FT SEQUENCE 303 AA: 33502 MW: 839D6CID CRC32;

US02_HSVK Length: 303 February 14, 2000 08:02 Type: P Check: 5996 ..

1 MGVLIVIVT VVDRHKLIPN SSIDVDGHLW EFLSRQCVL ASEPLGIPIV

51 VRSDILYRFS SSILTIPIKAC RPIYTRGDT AIALDRNGVY YHEDRMGYSI

101 EMLSVLSGYN HLNSSLITNQ PYHLWVGLAA DLCKPFDLI PGPKRMVYAE

151 IADEFKSQW PPVYCGKLEF TIPTVTEHN HPLKLAAG EDYVVGECGF

201 SKHSSNSLVH PPTVNRVYA VVDPARLEI PAFGRPLPR RPSEGMRAP

251 RRRSRAPAPA RSTAAATPP RPPGRAPAA RRAQDVYME RLWVGFGRT

301 STR

11AA_SEQUENCE 1.0 STANDARD; PRT: 384 AA.

ID VANS_ENTFC

AC 006240;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE SENSOR PROTEIN VANS (EC 2.7.3.-) (VANCOMYCIN RESISTANCE PROTEIN VANS)

DE (VANCOMYCIN HISTIDINE PROTEIN KINASE).

GN VANS.

OS Enterococcus faecium (Streptococcus faecium).

OC Bacterioidetes; Bacillus/Clostridium group; Enterococcaceae;

OC Enterococcus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BM4147; TRANSPOSON-TN1546;

RX MEDLINE: 93106944

RA ARTHUR M., MOLINAS C., DEPAIDIE F., COUVALLIN P.;

RT "Characterization of TN1546, a Tn3-related transposon conferring

RT glycopeptide resistance of TN1546, a Tn3-related transposon conferring

RT precursors in Enterococcus faecium BM4147.";

RL J. Bacteriol. 175:117-127(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN-BM4147;

RX MEDLINE: 92210502.

RA ARTHUR M., MOLINAS C., COUVALLIN P.;

RT "The Vans-Vanr two-component regulatory system controls synthesis of

RT desipeptide peptidoglycan precursors in Enterococcus faecium

RT BM4147.";

RL J. Bacteriol. 174:2582-2591(1992).

CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM VANS/VANR.

CC ACTIVATES THE TRANSCRIPTION OF VANH, VANM AND VANX IN RESPONSE TO

CC VANCOMYCIN WHICH RESULTS IN VANCOMYCIN RESISTANCE. VANS ACTIVATES

CC VANR BY PHOSPHORYLATION.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSOR TRANSDUCTION HISTIDINE

CC KINASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M97297; AAA65954.1; -

DR EMBL: M68510; AAA24788.1; -

DR PIR: PF00512; signal; 1.

RT Sensory transduction; Transferrase; Kinase; Transmembrane; Cell wall;

RT Antibiotic resistance; Phosphorylation; Plasmid.

FT TRANSMEM 21 41

FT TRANSMEM 76 96

FT MOD.RES 164 164

FT SEQUENCE 384 AA: 43915 MW: 89E01F7 CRC32;

VANS_ENTFC Length: 384 February 14, 2000 08:02 Type: P Check: 3601 ..

1 MYIKLNKNK DYSLERKLY MYVAIVYA IVEFLYIRSM IKGKLDWIL

51 SIENKDYDN HLDAMKLYOY SIRNNDIFI YVAIVISILI LCRVMSKFA

101 KYFDELINGI DVLQONEDKQ IELSAEMOVN EQKNTLKRI LERQEDAKL

151 AQRKNDVYM VLANDIKTPL TSIIIGYLSL DEARDMPVDQ KAKYVHITID

201 KAYREQLID EFFEITRYNL QTITLTKTHI DLYTMLVOMT DEYIPOLSAH

251 GKQAVIHAPE DLYVSGDPK LARVENNIUK NAAVSEDSN IIDITAGLSG

301 DVYSIEFKNT GSIPKDLAA IEFKYRLDN ARSSDTGGAG LGALAIKEII

351 VOHQGIYAE SNDNYTTFV ELPAMPDLVD KRRS

11AA_SEQUENCE 1.0 STANDARD; PRT: 383 AA.

ID VES2_HPV03

AC P36778;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE REGULATORY PROTEIN E2.

GN E2.

OS Human papillomavirus type 3.

OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomaviruses.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 94265501.
 RA DELIUS H.; HOFMANN B.;
 RT "Primer-directed sequencing of human papillomavirus types.";
 CC Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACGNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X74462; CAA52472.1; -
 CC PIR: S36552; S36552.
 CC DR HSSP: P17383; IDHM.
 CC DR PFM: PF00508; E2_N; 1.
 CC DR PFM: PF00511; E2_C; 1.
 CC KW Early protein; transcription regulation; Activator; DNA-binding;
 CC Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 CC
 CC SEQUENCE 383 AA; 44246 MW; CFBIDEA CRC32;
 CC
 CC VE2_HPV03 Length: 383 February 14, 2000 08:02 Type: P Check: 4819 ..
 CC
 CC 1 METLANRLDV CODKLELEYE KSDKLEDOI MHQMLRLQO ALYKARECG
 CC 51 LTHIGHQVVP PLSTVKAKAR SAIEVHVSLO QLOHSAHAD PWTLRDSRE
 CC 101 MMDTVPRKGM KRGGLTVEVR YDGDENKAMC YVQWRELIYO NYTDDNWKV
 CC 151 AGVSHSEGLY VMHEGQKTFY VKFKDARVY GDTGWDVHV GGYVHHDSF
 CC 201 DVSSTREIR APGLYACTT QAPTAQVGA SEGPRQKRR LETVYGRQOQ
 CC 251 QOQOQOQOQO HQTAPQOTT ERANQPLDTD RTRDDTTCR HPIGHRSDPD
 CC 301 CYPVHLRQD PNCLECFRFR LNKGNKLYS RTSSWRMSC ESENGCAVYT
 CC 351 IWTSTYGORE AFLSTVYKVP GIOVILGHMS MFT
 CC
 CC !!AA-SEQUENCE 1.0 STANDARD: PRT: 376 AA.
 CC ID VE2_HPV10
 CC AC P36781;
 CC DT 01-JUN-1994 (Rel. 29, Created)
 CC DT 01-JUN-1994 (Rel. 29, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE REGULATORY PROTEIN E2.
 CC GN E2.
 CC OS Human papillomavirus type 10.
 CC OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 CC [1]
 CC RX MEDLINE: 94265501.
 CC RA DELIUS H.; HOFMANN B.;
 CC RT "Primer-directed sequencing of human papillomavirus types.";
 CC CC Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACGNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA

CC REPLICATION.
 CC
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X74465; CAA52492.1; -
 CC PIR: S36535; S36535.
 CC DR HSSP: P17383; IDHM.
 CC DR PFM: PF00508; E2_N; 1.
 CC DR PFM: PF00511; E2_C; 1.
 CC KW Early protein; transcription regulation; Activator; DNA-binding;
 CC Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 CC
 CC SEQUENCE 376 AA; 43003 MW; 799683D3 CRC32;
 CC
 CC VE2_HPV10 Length: 376 February 14, 2000 08:02 Type: P Check: 273 ..
 CC
 CC 1 METLANRLDA CODKMLEEY KSDKLEDOI THWHLRLVEN ALYKARECG
 CC 51 LTHIGHQVVP PLSTVKAKAR NAEVHVALQ QLOESAVAH PWTLRDSRE
 CC 101 MMDTAPKGM KRGGLTVEVR YDGDENKAMC YVQWRELIYO NYSDDRWKV
 CC 151 PGKSYEBLY YTHENNNTY VNFKDACYV GETGKEVAV GGYVHHDAF
 CC 201 DVSSTREIS TPGVCTSTNT TPASTQAVG ASGEPQKQO RLEAVDGOHQ
 CC 251 QOQOQKSDT QKAERAGQ VSDRTRLCD TRSAHPVRHP SDPPCAPYIH
 CC 301 LRQDPNSLKC FRYRLHGRK KLYSRSSSTW RMSCESENQA AFVILWYSTD
 CC 351 TORTEFLNVV KYPPGIOVIL GYMSIF
 CC
 CC !!AA-SEQUENCE 1.0 STANDARD: PRT: 365 AA.
 CC ID VE2_HPV18
 CC AC P06790;
 CC DT 01-JAN-1988 (Rel. 06, Created)
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE REGULATORY PROTEIN E2.
 CC GN E2.
 CC OS Human papillomavirus type 18.
 CC OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 CC [1]
 CC RX MEDLINE: 93197132.
 CC RA MEISSNER J.;
 CC RT "Tagi is a single cut enzyme for HPV-18.";
 CC CC Nucleic Acids Res. 21:1041-1041(1993).
 CC CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACGNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05015; CAA28667.1; ALT_SEQ.
DR PIR: D6251; W2RL18.
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
KW Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 365 AA; 41294 MW; 1871C1FE CRC32;
VE2_HPV18 Length: 365 February 14, 2000 08:02 Type: P Check: 5006 ..
1 MOPKELISE RLSCVODKII DHYENDSKDI DSOIQWOLI RMENALFFAA
51 REHGIOTLNH QVPAVYNISK SKAKHAIELQ MALQGLAOSA YKTEDWTLOD
101 TCEELNTEP THCFKGGQI VQYFPGNKO NCMYVANDS VYMTDAGTW
151 DKATCVSHR GLYVKEGYN TFYIEKSEC EKYGTGTWE VHFGNVYDC
201 NDSMSTSD TSVATQLYKQ LQHTSPYSS TVSVGTATY GQTSATRPQ
251 HGLAKKHC GVPNPLGAA TPFGNKKRRK LCGNTTPII HLGDRNSLK
301 CLRYRKHS DHYRDISSTW HMGAGNENK GILTYTASE TQTKFLNTV
351 AIPDSQIIV GYKTM
11AA SEQUENCE 1.0 STANDARD: PRT: 376 AA.
ID VE2_HPV28
AC P50771;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
GN REGULATORY PROTEIN E2.
OS Human Papillomavirus type 28.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN
RP SEQUENCE FROM N.A.
RA DELIUS H.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31783; AAA79425.1; ..
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
SQ SEQUENCE 376 AA; 43234 MW; 466EDC47 CRC32;
VE2_HPV28 Length: 376 February 14, 2000 08:02 Type: P Check: 1816 ..
1 METLANRLDV CODKMLELYE KDSNKLDDQI MHMOLMRVEN ALLYKARECG
51 LTHIGHQVVP PLSTVKAKAR SAIEVHALL QLOESAAYAD SWTLRDSRE
101 MMDIVPKCW KRGVTVFVR YDGDETRSMC YVHWRDIFTO NYSDKWKVY

151 AGHVSIEGLY YIHGEQTFY VKFKDAVYV GETGKEVHV GKGVIIHHAF
201 DPVSTREIP AAGPLTSDT TKASTETSVG ATEGPOOKRO RLETLNMEQO
251 QROYPQTPST QTERASOPL DVTRSDODT TCPYVGHPS DPGCAPVHL
301 KGDNCNCKCF RYRLHKGKRR LYCKTSSIWR WSCSENOAA FVLIWYTSYS
351 ORNEFLSTVK VPPGIQVILG HSMFV
11AA SEQUENCE 1.0 STANDARD: PRT: 391 AA.
ID VE2_HPV2A
AC P25482;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 13-JUL-1998 (Rel. 36, Last annotation update)
GN REGULATORY PROTEIN E2.
OS Human Papillomavirus type 2a.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN
RP SEQUENCE FROM N.A.
RA HIRSCH-BEHNAM A.; DELIUS H.; DE VILLIERS E.M.;
RT "A comparative sequence analysis of two human papillomavirus (HPV)
RT types 2a and 57.";
RL Virus Res. 18:81-98(1990).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACGNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55964; -; NOT_ANNOTATED_CDS.
DR PIR: S15617; S15617.
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
KW Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 391 AA; 43233 MW; B3DE7AF6 CRC32;
VE2_HPV2A Length: 391 February 14, 2000 08:02 Type: P Check: 734 ..
1 METLANRDA COETILELYE KDSNKLDDQI KHMVQVLEN VMLFKARECG
51 MRRVGTAVP ALIYSKAKAC QALIEVQALQ TIMQSAVSTE AMTLRDTCLE
101 MMDAPPKCW KKKGOSVYK FDGSSDRDNI YTSMGFIYVQ DITLDSMHV
151 PGQVDELGLY YHADVAVNY VDFGTESLTY GYGTGTEVHV AGVYIHHITA
201 SVSSTQASAS DDEPLSPIRT AVSPVAPVYA ASASTAGR AAPPTOALCS
251 AOAPTSPPAK KQRYVGOQH PPDDSTIRYV EGEVECYNKR SIDSKRDP
301 RMGHGDTDSV PVILRGDAN CLKCFRYRVQ KHKDVLARYV SSTHWAGGN
351 GDKTAFVTLW YTSVEQRTTF LTRVSIKGL IALPGWSAF V

RP SEQUENCE FROM N.A.
 RX MEDLINE: 91303675.
 RA LINGU O., CRUM C.P., SILVERSTEIN S.J.:
 RT "Biologic properties and nucleotide sequence analysis of human
 RT papillomavirus type 51."
 RL J. Virol. 65:4215-4225(1991).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACGNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2E'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M62877; -. NOT_ANNOTATED.CDS.
 DR PIR: B40415; M2ML51.
 DR HSSP: P17383; 1DHM.
 DR PFAM: PF00508; E2.N; 1.
 DR PFAM: PF00511; E2.C; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 358 AA: 40908 MW: E0C694E CRC32:
 VE2_HP51 Length: 358 February 14, 2000 08:02 Type: P Check: 1816 ..
 1 METICHLNV COEKILDCYE LDSKLDVQI NWTLRLRYEA AMFYAREBN
 51 LRTINHOVVP ATTYSKQKAC QALEMHALQ SLNKSIDYNE PPTMAETCE
 101 LMCVAPROCF KKGITIVYI FDGNKNDAM YTSMKFIYI DNDKRVKING
 151 NVDYTGIIYT VNSKREYVO FKDEAKIYA QQWEVYMGST VITCEYVSS
 201 TCSDALSTTT TVEOLSNPT TNPITTCVGA KEAQTOORR OHLTEPDSST
 251 ISPLSYDNTN NOIHCSSGST NTGCHOSATQ TAFIVHLKGD TNCLKCFRYR
 301 FTKKGLYKN VSTWHTSN TKTGIVTIVE DSAHQRETFI KTIKPPSVY
 351 LSLGIMTL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 368 AA.
 ID VE2_HP52
 AC P36796;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 13-JUL-1998 (Rel. 36, Last annotation update)
 DE REGULATORY PROTEIN E2.
 OS Human papillomavirus type 52.
 NC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94265501.
 RA DELIUS H., HOFMANN B.:
 RT "Primer-directed sequencing of human papillomavirus types."
 RT Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACGNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2E'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS

CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X74481; CA52558.1; -.
 DR PIR: S36576; S36576.
 DR HSSP: P17383; 1DHM.
 DR PFAM: PF00508; E2.N; 1.
 DR PFAM: PF00511; E2.C; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 368 AA: 41739 MW: 918AE9D9 CRC32:
 VE2_HP52 Length: 368 February 14, 2000 08:02 Type: P Check: 4491 ..
 1 MESIPARLNA VOEKILDCYE ADSNDLNAQI EHMKILTRMCC VLFYKAKELG
 51 ITHIGHQVVP PPAVSKAKAC QALELOLAL ALNKTQYSTD GWTLOQTSLE
 101 MMAEPQKYE KKHGYTIVQ YDNDKNTMD YTNMEIYLL GECECTIVEG
 151 QVDYGGIYV CGGEKIYVVK FSDNAQYCV TGWEVHWGG QVIVCPASVS
 201 SNEVSTTEA VALCTETSST SAVSVGAKD HLPQPKRRR PDVDSNTRK
 251 YPNNLRGQO SYDSTRGLV TATECTNKR VAHTICTAPI IHLKGDNSL
 301 KCLRIRKTH KSLYVOISST WHWTSNCTN NNLGIVTITY SDETRQOFL
 351 KTVKIRPTVO VLOGVMSL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 317 AA.
 ID VEA_BPP22
 AC Q03544;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE EAA PROTEIN.
 GN Bacteriophage P22.
 OS Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
 NC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94018622.
 RA WILF D.L., HO Y.S., POWERS S., ROSENBERG M.:
 RT "The int genes of bacteriophages P22 and lambda are regulated by
 RT different mechanisms." J. Virol. 67:261-271(1993).
 RL MOL. Microbiol. 9:261-271(1993).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L06296; AAC18882.1; -.
 DR PIR: S35282; S35282.
 SQ SEQUENCE 317 AA: 35719 MW: 05933D59 CRC32:
 VEA_BPP22 Length: 317 February 14, 2000 08:02 Type: P Check: 7758 ..

1 MTIIRKIE LFVKSPLNG LTRGEOMELA RIALASDAE TVRINKFSG
 51 TCVTLEOQPN AADDVAVYIP LYAAPVPER ERIRREHAEM SOKTFGDVGP
 101 VGPLKHLKSE ALFAAADPSD PLEKADMOFL LMDAQRKNGI SDEFITRAAI
 151 EKLEINKRKQ WPEPKDCEPR LHIKEPESV VPECPALP YAOVKAVADL
 201 YALCMOSGEV VTYTPDEKA TWINNYSGT CVOETVKLER LQPALAGNSP
 251 VIPGWISCS ERMNDDESK PLAIFTKCL GQGMFVATYD DQGFEDYWG
 301 MEIGVSHWM QLPDPPL
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 177 AA.
 ID VGG_BP64
 AC P03644;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MAJOR SPIKE PROTEIN (G PROTEIN) (GPG).
 GN Bacteriophage G4.
 OS Bacteriophage G4.
 CC Viruses; ssDNA viruses; Microviridae; Microviridae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 79053264.
 RA GOSDON G.N., BARRELL B.G., STADEN R., FIDDES J.C.;
 RL "Nucleotide sequence of bacteriophage G4 DNA.";
 RL Nature 276:236-247(1978).
 RN [2]
 RP SEQUENCE OF 1-81 FROM N.A.
 RA MEDLINE: 78248733.
 RA SIMS J., DRESSLER D.;
 RL "Site-specific initiation of a DNA fragment: nucleotide sequence of
 the bacteriophage G4 negative-strand initiation site.";
 RL Proc. Natl. Acad. Sci. U.S.A. 75:3094-3098(1978).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RA MEDLINE: 96217890.
 RA MCKENNA R., BOWMAN B.R., IIAG L.L., ROSSMANN M.G., FANE B.A.;
 RL "Atomic structure of the degraded procapsid particle of the
 bacteriophage G4: induced structural changes in the presence of
 calcium ions and functional implications.";
 RL J. Mol. Biol. 256:736-750(1996).
 CC -1- FUNCTION: MAJOR SPIKE COMPONENT. INVOLVED IN THE ATTACHMENT TO
 THE BACTERIAL HOST.
 CC -1- SUBUNIT: THE VIRION IS COMPOSED OF 60 COPIES EACH OF THE F, G, AND
 J PROTEINS, AND 12 COPIES OF THE H PROTEIN. THERE ARE 12 SPIKES
 WHICH ARE EACH COMPOSED OF 5 G AND ONE H PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation in
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J02454; AAA32324.1;
 DR EMBL: V00657; CA24020.1;
 DR EMBL: M25080; AAA32328.1;
 DR PIR: A04252; ZGBPG4.
 DR PDB: 1GFF; 03-APR-96.
 KW Coat protein; 3D-structure.
 FT CONFLICT 27 A -> T (IN REF. 2).
 SO SEQUENCE 177 AA; 18820 MW; 700ABE43 CRC32;
 VGG_BP64 Length: 177 February 14, 2000 08:02 Type: P Check: 4484

1 MEOKTISKIN APINSTOLAA TKTPAVAPY LSVPLUSST ILLNATTAV
 51 THSGICLHVY RIDETNPINH HALSTAGSIS NVPADMIFA IREVDAGVY

101 PFAVVALYDV YPIETFNCK AISEKDAVTI DSHPTVGNQD VYAGIMLSN
 151 AMTASTISGV LSVQNVNREA TVLOPLK
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 372 AA.
 ID VGLI_HSV2H
 AC P13261;
 DT 01-JAN-1980 (Rel. 13, Created)
 DT 01-JAN-1980 (Rel. 13, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE GLYCOPROTEIN 1.
 GN G1 OR US7.
 OS Herpes simplex virus (type 2 / strain HG52).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Simplexvirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 87111457.
 RA MCGEOCH D.J., MOSS H.W.M., MCNAB D., FRAME M.C.;
 RL "DNA sequence and genetic content of the HindIII 1 region in the
 RL short unique component of the herpes simplex virus type 2 genome:
 RL identification of the gene encoding glycoprotein G, and evolutionary
 RL comparisons.";
 RL J. Gen. Virol. 68:19-38(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA DOLAN A.;
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
 CC 2 'GH' GB 'CG' GB 'GI' AND GE
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS 1, TO VZV GP1V,
 CC AND TO PRV GP63.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X04798; CA28485.1;
 DR EMBL: Z86099; CAB06714.1;
 DR PIR: F43674; F43674.
 KW GLYCOPROTEIN.
 FT CARBOHYD 156 156 POTENTIAL.
 FT CARBOHYD 169 169 POTENTIAL.
 FT CARBOHYD 175 175 POTENTIAL.
 FT CARBOHYD 243 243 POTENTIAL.
 SO SEQUENCE 372 AA; 39558 MW; E7FA7C46 CRC32;
 VGLI_HSV2H Length: 372 February 14, 2000 08:02 Type: P Check: 501
 1 MPGSLDGLA ILGLWCAG LGVAGPYSL VSDSLVDAGA VGPOGFVED
 51 LRVGEHLHFV GACVPHNTNY DGIETLFRYP LGNHCRVYH VYTLACPRR
 101 PAVAFICRS THRAHPAYP TLEIGLAROP LNRVATATD YAGLYLFRW
 151 VGSATNSRF VLGVALSANG TTVYNGSDYG SCDPAOLPFS APRIGPSSVY
 201 TPGASRPTTP RTTPPSRPR DPTAPADGTG TPAASGEIA PPNSTRASAE
 251 SRRLTVAGV IQAIPASII AFVFLGSCIC FHRRCORRYR RPRGOIYNPG
 301 GVSCAVNEAA MARLGAELRS HPNTPPKRRR RSSSTTMS LTSIAESEP
 351 GPVVLVSYP RPRSGTAPQ EV
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 366 AA.
 ID VGLI_HSV2A
 AC Q01017;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GLYCOPROTEIN M.
 GN 39.
 OS Herpesvirus salmaiti (strain 11).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 RN [1]
 RP MEDLINE: 92333688.
 RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
 RA NEWMAN C., WITTMANN S., CRAIXON M.A., COLEMAN H., FLECKENSTEIN B.,
 RA HONESS R.W.;
 RT "Primary structure of the herpesvirus salmaiti genome."
 RT J. Virol. 66:5047-5058(1992).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN M.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: X61346; CAA5662.1; -
 DR PIR: A36810; QOBENS.
 DR PIR: P01528; Herpes-glycop: 1.
 KW Transmembrane; Glycoprotein.
 FT TRANSMEM 18 34 POTENTIAL.
 FT TRANSMEM 87 103 POTENTIAL.
 FT TRANSMEM 152 168 POTENTIAL.
 FT TRANSMEM 214 230 POTENTIAL.
 FT TRANSMEM 236 252 POTENTIAL.
 FT TRANSMEM 274 290 POTENTIAL.
 FT TRANSMEM 305 321 POTENTIAL.
 FT DOMAIN 343 351 ARG/LYS-RICH (BASIC).
 FT CARBOHYD 57 57 POTENTIAL.
 SQ SEQUENCE 366 AA: 42182 MW: 14D9C2B CRC32;
 VGLM_HSVSA Length: 366 February 14, 2000 08:02 Type: P Check: 7071 ..
 1 MKKASRSDTF MLRTWIOQLV LFVIMFMSA ILPIASVVG LGFCYPPNL
 51 VDSYSLNLT RNAKHITPT LFLAPLFLV YITWSVYDL ASAIYYVGA
 101 LAIQARKTH LMSITLQTM INLVGSHML FIGIAMMTL QLFHYLSYK
 151 HWMLAFIYF LHFCLSYMHT LSLVSRNSPK WSYLMECHI PKOSILSTIL
 201 DYKPLCVNM YLSLALML VESLGFMAI GNSFTLVSD TVLASINLIYF
 251 VLTFWYMT EMLFDYKTL QFGYLVGSV GSILILPVL RYEAWEVSAN
 301 LKHTVAVNIA MIPMCVIM MRLFRYSQ VPKPENSYP LKPKRRKQ
 351 KODQOLIME TSDEEL
 11AA_SEQUENCE 1.0
 ID VIRB_AGR5 STANDARD: PRT: 237 AA.
 AC P17798;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE VIRB8 PROTEIN.
 GN Agrobacterium tumefaciens.
 OS Plasmid pTiC58.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 90318324.
 RA KUDAU G.A., DE VOS G., OMEN J., MCCAFFREY G., ZAMBRSKI P.;
 RT "The vir operon of Agrobacterium tumefaciens pTiC58 encodes 11 open
 RT reading frames";
 RT Mol. Gen. Genet. 221:256-266(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9031800.
 RA ROZANSKY P.M., POWELL B.S., SHTRASU K., LIN T.-S., MOREL P.,
 RA ZYRINK E.M., STECK T.R., KADO C.I.;
 RT "Molecular characterization of the vir region of Agrobacterium
 RT tumefaciens: complete nucleotide sequence and gene organization of
 RT the 28.63-kbp region cloned as a single unit."
 RT Plasmid 23:85-106(1990).
 CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
 CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
 CC TRANSFER TO PLANT CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: X53264; CAA37361.1; -
 DR EMBL: J03320; AAA91596.1; -
 DR PIR: S12348; B8AG58.
 KW Crown gall tumor; Plasmid.
 FT CONFLICT 21 22 MISSING (IN REF. 1).
 FT CONFLICT 128 129 SA -> TR (IN REF. 2).
 SQ SEQUENCE 237 AA: 26294 MW: CD671958 CRC32;
 VIRB_AGR5 Length: 237 February 14, 2000 08:02 Type: P Check: 5994 ..
 1 MKGSYVALV ARETLAEHYK EYEAFOIARA KSARRLSKVI AAVAITAVIG
 51 NVQAQFTAT MVLIRLVPV YLMIRPDGV DSEVSYSRLP AIOEAVNVA
 101 SLMEYVRLRE SYDADTAQYA YDLVSNFSAP MVRQNYOOFF NYPNPTSPQV
 151 ILGKHGRLEV EHTASNDVTP GVOQIRYKRT LIYDGKMPMA STWTAIVRYE
 201 KVTSLPGRLR LTNPGGLVVT SYOTSEDTVS NAGHSEP
 11AA_SEQUENCE 1.0
 ID VIRB_AGR6 STANDARD: PRT: 230 AA.
 AC P09781;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VIRB8 PROTEIN.
 GN Agrobacterium tumefaciens.
 OS Plasmid pTiA6.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88186901.
 RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
 RA NESTER E.W.;
 RT "Characterization of the virB operon from an Agrobacterium
 RT tumefaciens Ti plasmid."
 RT J. Biol. Chem. 263:5804-5814(1988).
 RN [2]
 RP REVISIONS.
 RX MEDLINE: 90170994.
 RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
 RA NESTER E.W.;
 RL J. Biol. Chem. 265:4768-4768(1990).

CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
CC TRANSFER TO PLANT CELLS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: J03216; -; NOT ANNOTATED_CDS.

DR PIR: I28621; B8AG6.

KW Crown gall tumor; Plasmid.

SEQUENCE 230 AA; 25362 MW; 44A8A66C CRC32;

VIB8_AGR76 Length: 230 February 14, 2000 08:02 Type: P Check: 2646 ..

1 MLVARESLAE HYKEVEAFOT ARAKSARRLS KIILAAVAIA ILGNVAQAF
51 IATWPLSLRL VPYILMIRAD GTVDSEVSIS RLPAQOEAV VNASLMEYR
101 LRESYDADTA QYAYDLVSNE SAPTVRODYO QFFNYPNSS POYLGRGR
151 VEVEHIASND VTPSTOQIRY KRLVVDGKM PVSTWTATV RYKVTSLPG
201 RLRLNPAGL VTSYOTSED TVSNVGGAP

11AA_SEQUENCE 1.0

ID VIB8_AGR79 STANDARD; PRT; 230 AA.

AC P05357;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 01-MAY-1992 (Rel. 22, Last annotation update)

DE VIRB8 PROTEIN.

GN Agrobacterium tumefaciens.

OS Plasmid pRi1595.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Agrobacterium.

OC [1]

RP MEDLINE; 88247765.

RA THOMPSON D.V., MELCHERS L.S., IDLER K.B., SHILPEROORT R.A.,
RA HOOYKAAS P.J.J.;

RT "Analysis of the complete nucleotide sequence of the Agrobacterium
RT tumefaciens virB operon."

RT Nucleic Acids Res. 16:4621-4636(1988).

CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
CC TRANSFER TO PLANT CELLS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X06826; CAA29978.1; AL1-INT.

DR PIR: S00784; B8AG55.

KW Crown gall tumor; Plasmid.

SEQUENCE 230 AA; 25408 MW; 870AF438 CRC32;

VIB8_AGR79 Length: 230 February 14, 2000 08:02 Type: P Check: 2826 ..

1 MLVARESLAE HYKEVEAFOT ARAKSARRLS KIILAAVAIA ILGNVAQAF
51 IATWPLSLRL VPYILMIRAD GTVDSEVSIS RLPAQOEAV VNASLMEYR
101 LRESYDADTA QYAYDLVSNE SAPTVRODYO QFFNYPNSS POYLGRGR

151 VEVEHIASND VTPSTOQIRY KRLVVDGKM PVSTWTATV RYKVTSLPG
201 RLRLNPAGL VTSYOTSED TVSNVGGAP

11AA_SEQUENCE 1.0

ID VIB8_AGR79 STANDARD; PRT; 147 AA.

AC P13461;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-JAN-1990 (Rel. 13, Last annotation update)

DE T-DNA BORDER ENDONUCLEASE VIRID1 (EC 3.1.1.-).

GN VIRID1.

OS Agrobacterium rhizogenes.

OS Plasmid pRi14b.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Agrobacterium.

OC [1]

RP MEDLINE; 89039712.

RA HIRAYAMA T., MURAKAWA T., OKAWA H., OKA A.;

RT "Organization and characterization of the virC genes from
RT Agrobacterium rhizogenes."

RT Mol. Gen. Genet. 213:229-237(1988).

CC -1- FUNCTION: TUMOR FORMATION BY A. TUMEFACIENS INVOLVES THE TRANSFER
CC INTO THE PLANT NUCLEAR GENOME. THE VIRD OPERON ENCODES A SITE-
CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X12867; CAA31350.1; -.

DR PIR: S06883; S06883.

KW Hydrolyase; Nuclease; Crown gall tumor; Plasmid;

KW T-DNA.

SEQUENCE 147 AA; 16005 MW; 414FDDA0 CRC32;

VID1_AGR76 Length: 147 February 14, 2000 08:02 Type: P Check: 2761 ..

1 MSOGRPTSS DIAINORVGA TVEGFRVST RLRSAYESF SHQARLLGS
51 DSAIRVAVR RIGGFLEIDA ETRHKEMALI LSIGTLSSNI ALLSAVAEN
101 PTMDELRLA ERINGESFA DDDGLRLSII SVSRRIIDGC SMKDSL

11AA_SEQUENCE 1.0

ID VIB8_AGR75 STANDARD; PRT; 147 AA.

AC P18551; P06520;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE T-DNA BORDER ENDONUCLEASE VIRID1 (EC 3.1.1.-).

GN VIRID1.

OS Agrobacterium tumefaciens.

OS Plasmid pRiC58.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Agrobacterium.

OC [1]

RP MEDLINE; 90301800.

RA ROGOWSKY P.M., POWELL B.S., SHIRASU K., LIN T.-S., MOREL P.,
RA ZYRAN E.M., STECK T.R., KAO C.I.;

RT "Molecular characterization of the vir regulon of Agrobacterium
RT tumefaciens: complete nucleotide sequence and gene organization of
RT the 28.63-kbp regulon cloned as a single unit."

RT Plasmid 23:85-106(1990).

```

RN PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE: 85190558
RA HAGIVA M., CLOST T.J., TAIT R.C., KADO C.I.;
RT "Identification of pTIC58 plasmid-encoded proteins for virulence in
RT Agrobacterium tumefaciens."
RT Proc. Natl. Acad. Sci. U.S.A. 82:2669-2673(1985).
RN [3]
RP SEQUENCE FROM N.A.
RA WANG K., HERRERA-ESTRELLA A.H., VAN MONTAGU M.M.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DBD databases.
CC -1- FUNCTION: TUMOR FORMATION BY A TUMEFACIENS INVOLVES THE TRANSFER
CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC INTO THE PLANT NUCLEAR GENOME. THE VIRID OPERON ENCODES A SITE-
CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03320; AAA91603.1;
DR EMBL: M11311; AAA98365.1; ALT_SEQ.
DR EMBL: M33673; AAA22110.1;
DR PIR: A22666; A22666.
DR PIR: S11838; S11838.
KM Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;
SQ T-DNA.
SEQUENCE 147 AA: 16130 MW: 7187146F CRC32:
VIDL_AGRT6 Length: 147 February 14, 2000 08:02 Type: P Check: 2312
1 MSQSRPTSS DIAYNORECV KYEGKRVST RLSAEYEF SHQARLLGLS
51 DSMAIRAVAR RIGGFLEIDA ETRHREMAIL QSIGTLSSNT AALLSAVAEN
101 PTMDLEALRA ERIAFKSFSA DDGLLRSL SVSRRRIDGC SLKDAL
11AA_SEQUENCE 1.0 STANDARD: PRT: 147 AA.
ID VIDL_AGRT6
AC P06667;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last annotation update)
DE T-DNA BORDER ENDONUCLEASE VIRID1 (EC 3.1.-.-).
GN VIRID1.
OS Agrobacterium tumefaciens.
OG Plasmid pTIC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 87028239.
RA YANOFSKI M.F., PORTER S.G., YOUNG C., ALBRIGHT L.M., GORDON M.P.,
RA NESTER E.W.;
RT "The virD operon of Agrobacterium tumefaciens encodes a site-specific
RT endonuclease."
RT Cell 47:471-477(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE: 88032822.
RA JAYASMAL R.K., VELUTHAMBI K., GELVIN S.B., SLIGHTOM J.L.;
RT "Double-stranded cleavage of T-DNA and generation of single-stranded
RT T-DNA molecules in Escherichia coli by a virD-encoded
RT border-specific endonuclease from Agrobacterium tumefaciens."
RL J. Bacteriol. 169:5035-5045(1987).
CC -1- FUNCTION: TUMOR FORMATION BY A TUMEFACIENS INVOLVES THE TRANSFER
CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC INTO THE PLANT NUCLEAR GENOME. THE VIRID OPERON ENCODES A SITE-

```

```

CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M41762; AAA98369.1;
DR EMBL: M1989; AAA22113.1;
DR PIR: A29826; A29826.
DR PIR: A25063; A25063.
KM Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;
SQ T-DNA.
SEQUENCE 147 AA: 16215 MW: 305D4C9C CRC32:
VIDL_AGRT6 Length: 147 February 14, 2000 08:02 Type: P Check: 4145
1 MSKHRTYSS ETRAIQHRSL NVGEFKVSA RLSAEYEF SYQARLLGLS
51 DSMAIRAVAR RIGGFLEIDA HTRKREMAIL QSIGTLSSNV SMLSAVAED
101 PRSDLEAVRD ERIAFGEAFA ALDGLLRSL SVSRRRIDGC SLKDAL
11AA_SEQUENCE 1.0 STANDARD: PRT: 374 AA.
ID VF1_NPVOP
AC O10330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VERY LATE EXPRESSION FACTOR 1.
GN VF1-1
OS Orygia pseudotsugata multicausid polyhedrosis virus (OpNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC Nucleopolyhedrovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 97271300.
RA ARENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orygia pseudotsugata multicausid multicausid nuclear
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
CC -1- FUNCTION: INVOLVED IN VERY LATE GENE ACTIVATION (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59079.1;
DR PIR: P00589; P00589.
KM Transcription regulation.
SQ SEQUENCE 374 AA: 43165 MW: C560A0FD CRC32:
VF1_NPVOP Length: 374 February 14, 2000 08:03 Type: P Check: 889
1 MDGLGVNET TENDMKARIQ SASREHVD LATDROCTP DEVANDSLMS
51 KYMPKPPAP TILSKYSRL IKIIFSLVE ADIQNPAYL DREDDSVEFO
101 HLVSPELC KRMLEIRSVT KETIQLTINE YTNANMLPEP KIPRMVMPR
151 DREKTIRES ENKFMKRNAI DTILNFDK IKIANSYVH DRGLIRGAIV
201 FCIMIGTMR INEAROLSYD DLNVLKRGK LRSDTIGLR KRSRKNLTNN

```

251 IKTKPLELAR EYARRPTVL QISKSTSTPF KDFRLLDEA GYEMERPRSN
301 MIRYLSNNL YNSGVPLQKV ARLMNHSPSA STKPYLNKNY FDESSDEES
351 GGNRRDSSTG SSANSSSLXY QTGD

11AA-SEQUENCE 1.0 STANDARD: PRT: 186 AA.
ID VMA2_HRSVA

AC P29792;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE MATRIX GLYCOPROTEIN M2.
GN M2.
OS Bovine respiratory syncytial virus (strain A51908) (BRS).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ZAMORA M., SAMAL S.K.;
RX MEDLINE: 92185490.
RT "Sequence analysis of M2 mRNA of bovine respiratory syncytial virus
RT obtained from an F-M2 distrophic mRNA suggests structural homology
RT with that of human respiratory syncytial virus.";
RL J. Gen. Virol. 73:737-741(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).

DR EMBL: M82816; AAA42805.1; -
DR PIR: J01482; WMNZBA.
KW Matrix protein; Envelope protein; Glycoprotein.
FT CARBOHYD 54 POTENTIAL.
FT CARBOHYD 89 POTENTIAL.
SQ SEQUENCE 186 AA; 21351 MW; CC402395 CRC32;

VMA2_HRSVA Length: 186 February 14, 2000 08:03 Type: P Check: 6337 ..

1 MSRRNPCKYE IRGHCLNGKR CHEFHNPFEM PPHALLVRON FULNKLKSM
51 DRNDITLSEI SGAALDRTE EYALGVIVL ESYLSISNNI TKOSACVAMS
101 KLAEINND IKRLRNKEVP TSPKIRIYNT VISYDSNKR NTKOTIHLK
151 RLPADVLKKT IKNTIDIHNE INGNNOGDIN VDEONE

11AA-SEQUENCE 1.0 STANDARD: PRT: 194 AA.
ID VMA2_HRSVA

AC P04545;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MATRIX GLYCOPROTEIN M2 (ENVELOPE-ASSOCIATED 22 KD PROTEIN).
GN M2.
OS Human respiratory syncytial virus (strain A2).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ELANGO N., SARAK M., VENKATESAN S.;
RX MEDLINE: 85237684.
RT "RNA sequence of three respiratory syncytial virus genes encoding
RT two nonstructural proteins and a 22k structural protein.";
RL J. Virol. 55:101-110(1985).

[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85135082.
RA COLLINS P.L., WERTZ G.W.;

RT "The envelope-associated 22k protein of human respiratory syncytial
RT virus: nucleotide sequence of the mRNA and a related
RT polytranscript.";
RL J. Virol. 54:65-71(1985).

[3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95266253.
RA CONNORS M., CROWE J.E. JR., FIRESTONE C.Y., MURPHY B.R., COLLINS P.L.;
RT "A cold-passaged, attenuated strain of human respiratory syncytial
RT virus contains mutations in the F and L genes.";
RL J. Virology 208:478-484(1995).

[4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97187925.
RA CROWE J.E. JR., FIRESTONE C.Y., WHITEHEAD S.S., COLLINS P.L.,
RA MURPHY B.R.;
RT "Acquisition of the ts phenotype by a chemically mutagenized cold-
RT passaged human respiratory syncytial virus vaccine candidate results
RT from the acquisition of a single mutation in the polymerase (L)
RT gene.";
RL Virus Genes 13:269-273(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).

DR EMBL: M11486; AAB59860.1; -
DR EMBL: U50362; AAB86655.1; -
DR EMBL: U50363; AAB86677.1; -
DR EMBL: U63644; AAC55971.1; -
DR PIR: A04034; WMNZ22.
KW Matrix protein; Envelope protein; Glycoprotein.
FT CARBOHYD 89 POTENTIAL.
FT CARBOHYD 191 POTENTIAL.
SQ SEQUENCE 194 AA; 22154 MW; 59973F8C CRC32;

VMA2_HRSVA Length: 194 February 14, 2000 08:03 Type: P Check: 9828 ..

1 MSRRNPCKYE IRGHCLNGKR CHEFHNPFEM PPHALLVRON FULNKLKSM
51 DKSIDITLSEI SGAALDRTE EYALGVIVL ESYIGSISNNI TKOSACVAMS
101 KLTELNSDD IKRLRDNDEL NSPKIRIYNT VISYIESNKR NTKOTIHLK
151 RLPADVLKKT IKNTLDIHS ITINPKEST VSDNDHAKN NDTT

11AA-SEQUENCE 1.0 STANDARD: PRT: 186 AA.
ID VMA2_ORSVW

AC O84132;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MATRIX GLYCOPROTEIN M2.
GN M2.
OS Ovine respiratory syncytial virus (strain MSU 83-1578) (ORSV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ALANSARI H.M., FORGIETER L.N.D.;
RX MEDLINE: 95088607.
RT "Molecular cloning and sequence analysis of the phosphoprotein,
RT nucleocapsid protein, matrix protein and 22k (M2) protein of the
RT ovine respiratory syncytial virus.";
RL J. Gen. Virol. 75:3597-3601(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U02510; AAA51645.1; -
KW Matrix protein; Envelope protein; Glycoprotein.
FT CARBOHYD 54
FT CARBOHYD 89
SQ SEQUENCE 186 AA; 21133 MW; CD7955FA CRC32;
VMA2_ORSW Length: 186 February 14, 2000 08:03 Type: P Check: 6559 ..
1 MSRRRRCAYE IRGHCLNGKR CHFSHYFEM PPHLLVLRON FEMKTLKEM
51 DSSNLTSEI SGAEEDRTE EYALGVIVL ESYGASVNI TRKSACVAMS
101 KLGEINSD IKGUNKEP TSPKRIYNT VISYDSNKR NPKOTHLK
151 RLPADVLRKT IKNTIDINE INVNPDSIG VNEONE
11AA_SEQUENCE 1.0 STANDARD; PRT; 106 AA.
VMEW_CVB
AC P37989;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 11.4 KD MEMBRANE PROTEIN (ORF 3).
OS Chrysanthemum virus B (CVB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92013948.
RA LEVAY K., ZAVRIEV S.;
RT "Nucleotide sequence and gene organization of the 3'-terminal region
RT of chrysanthemum virus B genomic RNA."
RL J. Gen. Virol. 72:2333-2337(1991).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEVIRUSES AND TO THE 14 KD
CC PROTEIN FROM BSMV RNA 2BETA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; S60150; CAB31868.1; -
DR PIR; J01248; J01248.
DR PFAM; PF01307; Plant_vir_prot; 1.
KW Transmembrane.
FT TRANSMEM 10
FT TRANSMEM 72
SQ SEQUENCE 106 AA; 11435 MW; CPA2CE4A CRC32;
VMEW_CVB Length: 106 February 14, 2000 08:03 Type: P Check: 566 ..
1 MFLPPDPT KVLVAIGL SIYASLTYS RNTLPQVGDH SHLLPGVY
51 KQGTIYVG GPRKNSLEG GFRLPYQPF LVLLSAIF LLSGRGHR
101 VCGQCH
11AA_SEQUENCE 1.0 STANDARD; PRT; 108 AA.
VMEW_LVX
AC P27331;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE HYPOTHETICAL 11.8 KD PROTEIN (ORF 3).
OS Lily virus X.

CC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90218039.
RA MEMELINK J., VAN DER VLOET C.I.M., LINTHORST H.J.M.,
RA DERKS A.F.L.M., ASJES C.J., BOL J.F.;
RT "Homologies between the genomes of a carlavirus (lily symptomless
RT virus) and a potexvirus (lily virus X) from lily plants."
RL J. Gen. Virol. 71:917-924(1990).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM OTHER POTEVIRUSES AND TO 12 KD
CC PROTEIN FROM CARLAVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X15342; CA33395.1; -
DR PFAM; PF01307; Plant_vir_prot; 1.
KW Transmembrane.
FT TRANSMEM 9
FT TRANSMEM 29
FT TRANSMEM 72
SQ SEQUENCE 108 AA; 11767 MW; 6EBDF03 CRC32;
VMEW_LVX Length: 108 February 14, 2000 08:03 Type: P Check: 8138 ..
1 MFLPPDPT KPFIAYVVG TLAAFLVLT RNTLPHTGDN LMSLPHGCTY
51 CDGTRIRYVG GPRSHVPEL PAKSMALITY VAILALHFS CLTRHVRHC
101 VICHTTSG
11AA_SEQUENCE 1.0 STANDARD; PRT; 109 AA.
VMEW_PVMR
AC P17527;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 12 KD MEMBRANE PROTEIN (ORF 3).
OS Potato virus M (Strain Russian) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89293091.
RA RUPASOV V.V., MOROZOV S.Y., KANYUKA K.V., ZAVRIEV S.K.;
RT "Partial nucleotide sequence of potato virus M RNA shows similarities
RT to potexviruses in gene arrangement and the encoded amino acid
RT sequences."
RL J. Gen. Virol. 70:1861-1869(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91116326.
RA ZAVRIEV S.K., KANYUKA K.V., LEVAY K.E.;
RT "The genome organization of potato virus M RNA."
RL J. Gen. Virol. 72:9-14(1991).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEVIRUSES AND TO THE 14 KD
CC PROTEIN FROM BSMV RNA 2BETA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D14449; BA03341.1; -
DR EMBL; X53062; CA37234.1; -
DR PIR; P00003; MWYF2.

DR PIR: S21603; S21603.
 DR PIR: C54333; C54333.
 DR PFAM: PF01307; Plant_vir_prot; 1.
 KW TRANSMEMBRANE.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 SO SEQUENCE 109 AA; 11907 MW; E58C570 CRC32;
 VMEH_PVR Length: 109 February 14, 2000 08:03 Type: P Check: 3415 ..
 1 MFLTPPDEFT KYLSAALGV SLATVWLLI RSTLPVGDH DNHPHGMY
 51 RDGTSVFEYN SPGRINSIA RKAPLAGPW AIYVLLVLLI WASHKLGPN
 101 CRACGSHT
 11AA_SEQUENCE 1.0
 ID VPR_SALDU STANDARD; PRT; 149 AA.
 AC P07579;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last annotation update)
 DE MAJOR NUCLEOCAPSID PROTEIN (P8 PROTEIN).
 GN P8.
 OS Bacteriophage phi-6.
 OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
 RN [1]
 RP MEDLINE: 86144085.
 RA MCGRAN I., MINICH L., FRANGIONE B.;
 RT "Nucleotide sequence of the small double-stranded RNA segment of
 bacteriophage phi 6: novel mechanism of natural translational
 control.";
 RL J. Virol. 58:142-151(1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M12921; AAA33358.1; -
 DR PIR: A23368; VMBPF6.
 KW Nucleocapsid.
 SO SEQUENCE 149 AA; 16004 MW; 23F29489 CRC32;
 VPR_BP6 Length: 149 February 14, 2000 08:03 Type: P Check: 6252 ..
 1 MLBPVRAAA VPAIESAIA TPGVSRRA AIGSKVSPS ILAAVKSMPV
 51 VAGITLAIQIG STGYDAVQOL LENBEVAVEM LKLSKRADE IOPDTIGLUG
 101 QYREBELVE DAARVVGWS NLIRLQALE LDIKITGLKM QLNMGYRS
 11AA_SEQUENCE 1.0
 ID VPR_SALDU STANDARD; PRT; 297 AA.
 AC P24417;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE VIRULENCE GENES TRANSCRIPTIONAL ACTIVATOR.
 GN VSDA OR SPVR.
 OS Salmonella dublin.
 OC Plasmid PSD12.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91251759.
 RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINNY D.;
 RT "Molecular analysis of the virulence locus of the Salmonella dublin
 RT

RT Plasmid PSD12.";
 RL Mol. Microbiol. 5:307-316(1991).
 CC
 CC -1- FUNCTION: POSITIVE REGULATOR FOR THE PLASMID-ENCODED VIRULENCE
 CC FACTORS SPVA, SPVB AND SPVC
 CC
 CC -1- SUBCELLULAR LOCATION: CYTOSOL; CYTOSOL;
 CC
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X56727; CAA40047.1; -
 DR PIR: S15213; S15213.
 DR PROSITE: P500044; HTH_LYSR_FAMILY; 1.
 DR PFAM: PF00126; HTH_1; 1.
 KW Plasmid; Virulence; Transcription regulation; Activator; DNA-binding.
 FT DNA_BIND 21 40 H-T-H MOTIF (POTENTIAL).
 SO SEQUENCE 297 AA; 33849 MW; DA84490B CRC32;
 VPR_SALDU Length: 297 February 14, 2000 08:03 Type: P Check: 7061 ..
 1 MFLINKKIK IFTIMETGS FSIATSVLXI TRTPLSRVIS GLERELKQRL
 51 FIRKNGTLIP TEFAQTIYRK VKSHYFELHA LEQELGPRG TKOLEIIFDE
 101 IYPSIAKNI ISALTISGOK TINMGAVNS QIIELOCOTN NCIVISARY
 151 FHRRESLVCT SVEGVAWFT PKKFLCGKP DINRLAGIPV LRECKAKNN
 201 LDTIHFEEQ TLGITNPFS FDNVLFSSL YRLOQGLAML LIPVRCRL
 251 GSTDHAIHI KVALCTSLY YPTKRETPD YRKAIKLIQ ELKOSTE
 11AA_SEQUENCE 1.0
 ID VPR_SALDU STANDARD; PRT; 297 AA.
 AC P13041;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VIRULENCE GENES TRANSCRIPTIONAL ACTIVATOR.
 GN MKAC OR SPVR OR VAGA.
 OS Salmonella typhimurium.
 OC Plasmid 96 kb virulence pex102.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1275 WILD TYPE;
 RX MEDLINE: 90016881.
 RA FOLINGER G.D., BAIRD G.D., WILLIAMSON C.M., LAX A.J.;
 RT "Nucleotide sequence of a plasmid gene involved in the virulence of
 RT salmonella.";
 RL Nucleic Acids Res. 17:7983-7983(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90060335.
 RA TAIRA S., RHEIN M.;
 RT "Molecular organization of genes constituting the virulence
 RT determinant on the salmonella typhimurium 96 kilobase pair plasmid.";
 RL FEBS Lett. 257:274-278(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92041614.
 RA CALDWELL A.L., GULIG P.A.;
 RT "The salmonella typhimurium virulence plasmid encodes a positive
 RT regulator of a plasmid-encoded virulence gene.";
 RL J. Bacteriol. 173:7176-7185(1991).
 RN [4]

RP SEQUENCE OF 1-10.
 RX MEDLINE: 91244158.
 RA TAIRA S., BAUMANN M., RIKKONEN P., SUKUPOLVI S., RHEIN M.;
 RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
 associated proteins of *Salmonella typhimurium*.";
 RT FEMS Microbiol. Lett. 61:319-323(1991).
 CC -1- FUNCTION: POSITIVE REGULATOR FOR THE PLASMID-ENCODED VIRULENCE
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X16111; CAA34244.1; .
 CC PIR: S06089; S06089.
 CC PIR: S06670; S06670.
 CC PROSITE: PS00044; HTH_LYSR_FAMILY: 1.
 CC PRAM: PF00126; HTH_1: 1.
 CC Plasmid: Virulence; Transcription regulation; Activator; DNA-binding.
 CC DNA BIND 21 40 H-T-H MOTIF (POTENTIAL).
 CC SEQUENCE 297 AA: 33835 MW; 765F2667 CRC32;
 SQ
 VRRP_SALTY Length: 297 February 14, 2000 08:03 Type: P Check: 7032 ..
 1 MDPLINKKK IFITLMEGTS FSATSVLYI TRTPLSRVIS DLERELKRL
 51 FIRKNGLLIP TEFQITTYRK VKSHYIFLHA LEOEIQPTGK TKOLEITIDE
 101 IYPSLEKNLI ISALITSGOK TNMGRAVNS QITEELCQTN NCIVISARNY
 151 FHRESLVGRT SVGGVMEFT PKKFLCGKP DINRLAGTVP LFHESAKNFN
 201 LDTYHEFEQ TLTITNPAFS FDNVDLFSL YRLQGLAML LIPVAVCRAL
 251 GLSTDHALHI KGVALCTSLY YPTKRRETPD YRKAIRLIDQ ELKOSTF
 11AA_SEQUENCE 1.0
 ID WN11_CHICK STANDARD: PRT: 354 AA.
 AC P49339;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95298011.
 RA TANDA N., OHUCHI H., YOSHIOKA H., NOJI S., NOHNO T.;
 RT "A chicken Wnt gene, Wnt-11, is involved in dermal development.";
 RT Biochem. Biophys. Res. Commun. 211:123-129(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96063018.
 RA TANDA N., KAWAKAMI Y., SAITO T., NOJI S., NOHNO T.;
 RT "Cloning and characterization of Wnt-4 and Wnt-11 cDNAs from chick
 RT embryo.";
 RT DNA Seq. 5:277-281(1995).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF DERMAL STRUCTURE,
 CC BOTH LIMB AND FEATHER BUDS. IS LIKELY TO SIGNAL OVER ONLY FEW CELL
 CC DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE DERMATOME. THE EXPRESSION

CC DOMAIN IS MUTUALLY EXCLUSIVE TO THE OTHER WNT GENES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D31901; BA06699.1; .
 CC PROSITE: PS00246; WNT1: 1.
 CC PRAM: PF00110; wnt: 1.
 CC Developmental protein: Glycoprotein; Signal.
 CC SIGNAL 1 24 POTENTIAL.
 CC CHAIN 25 354 WNT-11 PROTEIN.
 CC FT CAROHD 40 40 POTENTIAL.
 CC FT CAROHD 50 90 POTENTIAL.
 CC FT CAROHD 300 300 POTENTIAL.
 CC FT CAROHD 304 304 POTENTIAL.
 CC SEQUENCE 354 AA: 39507 MW; 14039075 CRC32;
 SQ
 WN11_CHICK Length: 354 February 14, 2000 08:03 Type: P Check: 4630 ..
 1 MKSPQFELFA AFLSLILQTG ICYGIKIAL SKTPSSLALN QTOCHQLEG
 51 LVYSQVQLCR SNLEMQTII QAAREVIKTC RTFESDMRWN CSSIELAPNY
 101 LIDLERTRE SAFYALSA AISHITARAC TTGDLPGCC GPITGETPGP
 151 GYRWGCAQDN LNYGLNGSK FSDAPMKMK SSGQAKRLMH LHNSEVRQV
 201 LKSLMKRK CHVSGGSCI KTCWKGLQEL RDIALDLNKL YLSATKVYHR
 251 PMGTRELYVP KDIDIRPYKE TELIYQSSP DECMKNKAVG SHGTQDRQN
 301 KISNGSDCD LMCGRGYNP YMDKYVERCH CKYHCCIVT CKKCERIVER
 351 YVCK
 11AA_SEQUENCE 1.0
 ID WN11_CHICK STANDARD: PRT: 354 AA.
 AC P51891;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Coturnix coturnix japonica (Japanese quail).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Coturnix.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97178987.
 RA EISENBERG C.A., GOURDIE R.G., EISENBERG L.M.;
 RT "Wnt-11 is expressed in early avian mesoderm and required for the
 RT differentiation of the quail mesoderm cell line QCE-6.";
 RT Development 124:2523-2536(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF DERMAL STRUCTURE,
 CC BOTH LIMB AND FEATHER BUDS. IS LIKELY TO SIGNAL OVER ONLY FEW CELL
 CC DIAMETERS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X97549; CAA66151.1; -
 DR PROSITE: PS00246; WNT1.1.
 CC Developmental protein; glycoprotein; signal.
 KM Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24
 FT CHAIN 25 354 WNT-11 PROTEIN.
 FT CARBOHYD 40 40 POTENTIAL.
 FT CARBOHYD 90 90 POTENTIAL.
 FT CARBOHYD 300 300 POTENTIAL.
 FT CARBOHYD 304 304 POTENTIAL.
 SQ SEQUENCE 354 AA; 39457 MW; 3FC9C358 CRC32;

WNT1.COTVA Length: 354 February 14, 2000 08:03 Type: P Check: 4030 ..

1 MRSPQFLA AFLSLIQTG ICYGIKIAL SKTPSALALN OTQCKOLEG
 51 LVYSQVLCR SNLEIMOTII QAAREVITKC RTFSDMRWN CSSIELAPNY
 101 LLDLEGTRE SAFYALSA AISHITARAC TTGDLPGCCG GPIDGEPGP
 151 GYRWGACADN LNYGLMGSK FSDAPMKMKK SSGQANKLMH LHNSEVGROY
 201 LRASLEMKCK CHGVSGSCI KTCWKGLQEL RDIALDLKKN YLSATKYVHR
 251 PMGTRKHLVP KDIDIRPYKE TELIYLQSSP DECMKNEKVG SHGTQDRQCN
 301 KTSNGSDSCD LMCCGRGYNP YMDKVERCH CKYHWCCTVT CRRCRTVER
 351 YVCK

11AA SEQUENCE 1.0
 ID WNT1_HUMAN STANDARD: PRT: 354 AA.
 AC 096014;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98434463.
 RA LAKO M., STRACHAN T., BULLEN P., WILSON D.I., ROBSON S.C., LINDSAY S.;
 RT "Isolation, characterization and embryonic expression of WNT1, a gene
 RT which maps to 11q13.5 and has possible roles in the development of
 RT skeleton, kidney and lung."
 RL Gene 219:101-110(1998).
 CC -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
 CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF
 CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Y13843; CAA74159.1; -
 DR EMBL: Y13844; CAA74159.1; JOINED.
 DR EMBL: Y13845; CAA74159.1; JOINED.
 DR EMBL: Y13846; CAA74159.1; JOINED.
 DR EMBL: Y13847; CAA74159.1; JOINED.

DR EMBL: Y12692; CAA73223.1; -
 DR MIM: 603699; -
 DR PROSITE: PS00246; WNT1.1.
 CC Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24
 FT CHAIN 25 354 WNT-11 PROTEIN.
 FT CARBOHYD 40 40 POTENTIAL.
 FT CARBOHYD 90 90 POTENTIAL.
 FT CARBOHYD 160 160 POTENTIAL.
 FT CARBOHYD 300 300 POTENTIAL.
 FT CARBOHYD 304 304 POTENTIAL.
 SQ SEQUENCE 354 AA; 39407 MW; A4F06B92 CRC32;

WNT1_HUMAN Length: 354 February 14, 2000 08:03 Type: P Check: 3073 ..

1 MRAPVCEA LFLALQTG VCYGIKIAL SKTPSALALN OTQCKOLEG
 51 LVSAVOLCR SNLEIMHTYV HAAREVMKAC RRAFDMRWN CSSIELAPNY
 101 LLDLEGTRE SAFYALSA TISHAIARAC TSGDLPGCCG GPVGEPPGP
 151 GNRWGRCADN LSYGLMGAK FSDAPMKYK TSGQANKLMH LHNSEVGROA
 201 LRASLEMKCK CHGVSGSCI RTCKWGLQEL QDVAADLTKR YLSATKYVHR
 251 PMGTRKHLVP KDIDIRPYKD WELVYLQSSP DFCMKNEKVG SHGTQDRQCN
 301 KTSNGSDSCD LMCCGRGYNP YTDKVERCH CKYHWCCTVT CRRCRTVER
 351 YVCK

11AA SEQUENCE 1.0
 ID WNT1_MOUSE STANDARD: PRT: 354 AA.
 AC P46615;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-SWISS.
 RX MEDLINE: 96028534.
 RA CHRISTIANSEN J.H., DENNIS C.L., WICKING C.A., MONKLEY S.J.,
 RA WILKINSON D.G., WAINWRIGHT B.J.;
 RT "Murine Wnt-11 and Wnt-12 have temporally and spatially restricted
 RT expression patterns during embryonic development."
 RL Mech. Dev. 51:341-350(1995).
 CC -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
 CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF
 CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X70800; CAA50070.1; -
 DR MGD: MG1:101948; WNT11.
 DR PROSITE: PS00246; WNT1.1.
 DR PRM: PF00110; Wnt1.1.
 KM Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24
 FT SIGNAL 24 POTENTIAL.

```

FT CHAIN 25 354 WNT-11 PROTEIN.
FT CARBOHYD 40 40 POTENTIAL.
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 160 160 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 304 304 POTENTIAL.
SQ SEQUENCE 354 AA: 39135 MW: 58C138B0 CRC32:

WNT1_MOUSE Length: 354 February 14, 2000 08:03 Type: P Check: 1689

1 MRAPQCEA LFLALHTG VCYGIKML SKTPALALN QTHCKQJEG
51 LVSQVQLCR SNLEKRTIV HAQAKAKC RRAFDKRN CSSILANV
101 LLDIERSTRE SAFVYALSA TISHTIARAC TSGDLPCCSC GPVGEPPGP
151 GNRNGGACADN LSYGLMGAK FSDAPMYKK TGSQANKLMR LHNSEVGRQA
201 LRASLETKCK CHGVSGGCSI RTCKWGLQEL QDVADLKT R YLSATKVVHR
251 PMGTRKHLVP KLDIRPVND SELVYLOSP DFCMKNEKYV SHGTODRCN
301 KTSNGSDSCD LMCGRGYNP YTDYVERCH CKYHMCQYVT CRCEHYER
351 YVCK

11AA_SEQUENCE 1.0
ID WNT4_DROME STANDARD: PRT: 389 AA.
AC P40589;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE WNT-4 PROTEIN PRECURSOR (DWMt-4).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganeae; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95171909.
RA GRABA Y., GIESLER K., ARAGNO D., LAURENTI P., MARIOL M.-C.,
RA BRENER H., SAGNER T., PRADL J.;
RT "DmWt-4, a novel Drosophila Wnt gene acts downstream of homeotic
RT complex genes in the visceral mesoderm."
RT Development 121:209-218(1995).
CC -1- FUNCTION: ACTS DOWNSTREAM OF HOMEOTIC COMPLEX GENES IN THE
CC -1- VISCERAL MESODERM AND IS REQUIRED FOR EMBRYONIC SEGMENTATION.
CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L25316; AAA67470.1;
CC FLYBASE: FBgn0010453; Wnt4.
CC PROSITE: PS00246; Wnt1; 1.
CC PFAM: PF00110; Wnt1; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 389 WNT-4 PROTEIN.
FT CARBOHYD 134 134 POTENTIAL.
FT CARBOHYD 269 269 POTENTIAL.
SQ SEQUENCE 389 AA: 43245 MW: 803759E3 CRC32:

WNT4_DROME Length: 389 February 14, 2000 08:03 Type: P Check: 8394

1 MGSHPHGHAL AGLAKGLITV AGGQGLPGNTL GYGSTMLNG GVGGAAGMGL

11AA_SEQUENCE 1.0
ID WNT4_HUMAN STANDARD: PRT: 153 AA.
AC Q15053;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0040.
GN KIAA0040.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE: 96051387;
RA NOMURA N., MIYAJIMA N., SAZUKA T., TANAKA A., KANARABAYASI Y.,
SATO S., NAGASE T., SEKI N., ISHIKAWA K.-I., TABATA S.;
RT Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.
RL DNA Res. 1:27-35(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA BIRD C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: D25539; BAA05022.1;
DR EMBL: 299715; CAB16870.1;
KM Hypothetical protein.
SQ SEQUENCE 153 AA; 17035 MW; 3D601FF8 CRC32;
Y040_HUMAN Length: 153 February 14, 2000 08:03 Type: P Check: 4815 ..
1 MARYHARYT TOPRANKPOTK CPFGGSGSGP RGFIDTYLA AMCPINALLT
51 ADGMPPTCL WATPHAKKE HSLHLMNVP KCVHMHVHT HTNSGKRYVG
101 KYLLIKMSL AMEYQGSTL STYTKMSHGK ALPDSITYQ FPNOQPHPT
151 SIP
!!AA-SEQUENCE 1.0
ID Y209_MYCGE STANDARD: PRT: 308 AA.
AC P47451;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN MG209.
GN MG209.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 96026346.
RA FRASER C.M., GOCARINE J.D., WHITE O., ADAMS M.D., CLATON R.A.,
RA FLEISCHMANN R.D., BOLT C.J., KERAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMAN J.L., WEIDMAN J.F., SMALL K.V., SANDOSKI M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOLT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium".
RT Science 270:397-403(1995).
RL [2]
RP SEQUENCE OF 125-243 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 94075230.
RA PETERSON S.N., HU P.-C., BOLT K.F., HUTCHISON C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing".
RL J. Bacteriol. 175:7918-7930(1993).
!!-SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: U39700; AAC71427.1;
DR EMBL: U02214; AAD12509.1;
DR TIGR: MG209.
DR PROSITE: PS01129; PST_RLU; 1.
DR PIRAM: P00849; YABO; 1.
DR PIRAM: P01479; S4; 1.
KM Hypothetical protein.
SQ SEQUENCE 308 AA; 34996 MW; DF6BECEP CRC32;
Y209_MYCGE Length: 308 February 14, 2000 08:03 Type: P Check: 590 ..
1 MKCFVYVTT KRUDSLASL LNSKRVYK LNMGOIKVN EKLTFKNSLI
51 VAKDVIVKE IHDETSDFI TSVEPYNKL EYLFEDKDIM VINKPSGLT
101 HPTFENEKAS LLAACIFHNN KNPVLYVRL DRDTSGAIVV CKNOASLNL
151 ONOLNRTLK RYVALVHP FNALGSINA PLARVNNKV MEKIQOTAKA
201 KQATKFKVI NQNEKAALIS LELLGRHQ IRVHLKFIQH PYNDPLVGI
251 KSEKDSYQG FLHANNICFI HPTLNKPMDF HAPLEPKFSI KIKSLNLSLT
301 DPLHYLFK
!!AA-SEQUENCE 1.0
ID Y209_MYCPN STANDARD: PRT: 309 AA.
AC P75485;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN MG209 HOMOLOG.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae".
RT Nucleic Acids Res. 24:4420-4449(1996).
RL !!-SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: AE000053; AAB96191.1;
DR PROSITE: PS01129; PST_RLU; 1.
DR PIRAM: P00849; YABO; 1.
DR PIRAM: P01479; S4; 1.
KM Hypothetical protein.
SQ SEQUENCE 309 AA; 35228 MW; 3375E3F5 CRC32;
Y209_MYCPN Length: 309 February 14, 2000 08:03 Type: P Check: 6533 ..
1 MEQTSVYTA QRDITFLATL LNLNRVYKAK LIYDGLSVN GKIKRKNML
51 VQPEBRVAVN WSELEFKVP VEQPYDPL DILXEDQIN VYKRNGLIS

```

101 HPTSFNESES ILGALFHCN HQPVFLVHRL DRDTSQVIML AKNOSLSLHL
151 OKOLOQRVMK RYLLALVHP LDISGTSISA PLERVGNKVK MKVGNSSSNK
201 AKNAFTFTV LNQNEKALI KCELLTGRT QIRVHLOFK HPVNDPLYG
251 LKSPQATVEYG OYLHAQOISF IHPTLNKEMG FEAQDKTFS DKLNLNLKI
301 ANSLXALFO

11AA_SEQUENCE 1.0
ID Y280_MYCGE STANDARD: PRT: 265 AA.
AC 01-OCT-1996 (Rel. 34, Created)
DC 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG280.
GN MG280.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FEISCHMANN R.D., BULL C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FORSMAN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U93707; AAC71502.1; AUT_INIT.
CC TIGR: MG280;
CC KM Hypothetical protein; Transmembrane.
CC FT TRANSMEM 12 32 POTENTIAL.
CC FT TRANSMEM 201 221 POTENTIAL.
CC SQ SEQUENCE 265 AA: 29566 MW: 5464E0FF CRC32:
Y280_MYCGE Length: 265 February 14, 2000 08:03 Type: P Check: 9511 ..

1 MVIFINFSK TILFFGMLV FVLLGFCIT ALYFRSTAN LVQARNSID
51 SRFSAKAFK NALANSANOF SKSSTNNLD QVKKDEQSL QVDEYKKNL
101 ESQNNLGNIS QEKIRLELAT KKDLENSIQ LDFKNNLKD NSTASSSSV
151 KKAQTADGVI SAVSEFSTQA OSIYSSYKI KNNIPSEOF NNYDYTMIT
201 IVAVSGMLA ILITTVESF LTKKRGLIR FSRFISTQL ADVHNDIIR
251 YPDLEEVIT ALDHD

11AA_SEQUENCE 1.0
ID Y373_TREPA STANDARD: PRT: 477 AA.
AC 083388;
DC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL PROTEIN TP0373.
GN TP0373.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE: 98332770.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.R., CHIDAMBARAM M., UTTERBACK T.,
RA McDONALD L., ARTACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete."
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE001216; AAC63358.1;
CC KM Hypothetical protein.
CC SQ SEQUENCE 477 AA: 52605 MW: 6763C3F8 CRC32:
Y373_TREPA Length: 477 February 14, 2000 08:03 Type: P Check: 4876 ..

1 MESSRQKLRP LLVHARSFG HFLVPRKPSG LLVAVSGAD SLALYNAHE
51 LADFQVCAC AVTDHSLRA QEGALDARFY RALCARFSPP LPFVQOISA
101 GAVHAKAKIR GROYGDARA LRKVFEDHYA ANCGAQQVLT AHTRDDQYET
151 LMRRLQOGAA ASALOGIRAA RGRVVRPLK VSRICEVDPL QTRVRRERD
201 ASNTCKRIYR NRIRRELTPA LDVAVLWRS GLDKTFPAGS AHSFCAAL
251 TMRRECSHRA WEVPRALGT RLMPNSDPL AAEPLRFL LQECVNLGV
301 SHRPVGALE RCAREGVRR IHVSGQLER AGAVYLFSCI HASDPAETK
351 KODAGSPSS EKQVSAIYV ARGAYPCAC GTLIVEVPA GVFCQAQDH
401 VGVGPFSPFP YIRTRTQDT ISIRGCHKGI RRMFSEWHP LSDRTVLPMI
451 EODGVLRALY GALTGYQNRV AERTPHE

11AA_SEQUENCE 1.0
ID Y425_MYCPN STANDARD: PRT: 450 AA.
AC P75172;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE RNA HELICASE MG425 HOMOLOG.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RT Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000023; AAB95867.1; -
DR PFAM: PF00270; DEAD.1.
DR PFAM: PF00271; Helicase.C.1.
KW Hypothetical protein; Helicase; ATP-binding; RNA-binding.
FT NP_BIND 47 54 ATP (POTENTIAL).
FT SITE 154 157 DEV BOX.
SQ SEQUENCE 450 AA; 51244 MW; 4D9D5B21 CRC32;

Y425_MYCPN Length: 450 February 14, 2000 08:03 Type: P Check: 2944

1 MDSFENELGV SPALIALTKD NNINOPTTIO QLAIPQFLQH QNLIHSPFG
51 TGTAVFGIP VETLTKKPS KGTOTLVYA PTELAQIK TPEINAKHT
101 HAKVSLIG IPIMQOLKOL ENQPEIVGCT MGRVADLER GVAKFEHLEH
151 LIIDEVDLML DRGFRKLEFD LLSRIKFEQ IAVYSASANE ETETAKQIT
201 KNGIFLAPE LKONAPEDPN KLIDQFVCL FSNRKKQALY SLVSQTRAKS
251 IIVFCOTKRL VDELICFLRK NDVKTYPHLG DKAQIRERN LKLFANTTAP
301 IYLVITDLIG RGIHVEGVDN VVNSACVNF ETYLHMGRT GRNNHSGSCI
351 TCTSHENKA FLKLEQVND KRISPLRPR LRLPIKCT QPKKGLSLD
401 SVQKIYVNR SNGTFKRVPL AGDLFSRNR QPERDMQNK LHSQMSNM

!!AA_SEQUENCE 1.0
Y464_MYCGE STANDARD; PRT; 385 AA.

AC P47702;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL PROTEIN MG464.
GN MG464.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 96026346;
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMAN R.D., BUTT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL Science 270:397-403(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: TO A M. CAPRICOLUM HOMOLOG.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U39728; AAC72484.1; -
DR EMBL: U39733; AAB01654.1; -
DR TIGR: MG464; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 39 59 POTENTIAL.

FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
SQ SEQUENCE 385 AA; 44207 MW; D341F781 CRC32;

Y464_MYCGE Length: 385 February 14, 2000 08:03 Type: P Check: 2836

1 MPILAKQINK ELKTFENPFM SAAVNEKNM WNKKEFSAI FIVIKYEIF
51 IFLIVLNG CQQLAQPT GTNOVLGSL EIGYKFGTG DYRDLSNN
101 GPFYFSDY TLAYGPFYM FWPAAQIVL PIMYATRVPL GSGVELGFNM
151 ILSLIVLL VRLITIVTL NSTLAEKN EVOGLAEIN AKYGALDLO
201 SKNRQLEIM SLKKNHKS SAAVOEVT LPFLIYRI VTTLPYKAI
251 ILNFMDLSK VPLTEFSNF TTGMPPIIF LVILPVOFL SQILPYWAS
301 KRNENAKHS OKSIBOLNT KMOLEFVY PAITAESNA GGVYWEFNA
351 LFTLOSYLE HVEIVYRREK RKQYSKLDL ILERE

!!AA_SEQUENCE 1.0
Y464_MYCPN STANDARD; PRT; 385 AA.

AC P75112;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG464 HOMOLOG.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885;
RA HIMELREICH R., HILBERT H., PLAGENS H., PIKUL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO A M. CAPRICOLUM HOMOLOG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000017; AAB95810.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
SQ SEQUENCE 385 AA; 44239 MW; F31AEB96 CRC32;

Y464_MYCPN Length: 385 February 14, 2000 08:03 Type: P Check: 5866

1 MPILINKKH ELKTFENPFM SAAVNEKNM LKNLKRAMCI IFKYLKVAIF
51 IFLIVLNG CQQLAQPT GTNOVLGSL EIGYKFGTG DYRDLSNN
101 GPFYFSDY TLAYGPFYM FWPAAQIVL PIMYATRVPL GSGVELGFNM
151 ILSLIVLL VRLITIVTL NSTLAEKN EVOGLAEIN AKYGALDLO

201 SKRNOMEM SLKKNHKS SASFOVEVT LPILITRI VTTLRIRAI
 251 ILFNMWLSK VPLTEISNF TSGWTFITF LIIVPOFI SOKLPIMAS
 301 KRNENAKAS OKSIEOLNKT KRMOLIFEV FAVITASAA GVCVYFELNA
 351 LEFTLOSILT HLFVKKRTR RLTYSKLEO MLERE

!!AA_SEQUENCE 1.0
 ID YAEK_RHISN STANDARD: PRT: 248 AA.
 AC P55434;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PUTATIVE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE YAEK (EC 1.-.-.-).
 GN YAEK.
 OS Rhizobium sp. (strain NGR234).
 OG Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97305956.
 RA FRIEBERG C.A., FELLAY R., BAIRDOCH A., BROUGHTON W.J., ROSENTHAL A.,
 RA PERRET X.: "Molecular basis of symbiosis between Rhizobium and legumes."
 RT Nature 387:394-401(1997).
 RL Nature 387:394-401(1997).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR). STRONG, TO BACTERIAL 'YDFG'.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000071; AAB91654.1; -
 DR HSSP: P14061; 1FDV.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 KM Hypothetical protein; Oxidoreductase; Plasmid.
 FT NP_BIND 8 32 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 147 147 BY SIMILARITY.
 SQ SEQUENCE 248 AA: 26679 MW: 47C5BF45 CRC32:

YAEK_RHISN Length: 248 February 14, 2000 08:03 Type: P Check: 5477 ..

1 MHPMREYA LYTGAASSGIG KATALELASA GLRYVALGRD RAALDEHST
 51 AGIVPVEVL SDVSEYVIGKI AGEKIDVLVN NAGLITASNS LVNLSDDID
 101 AMIDINIRSV EKLTRHVLKO MERRRGHIF FTGSSGGLAP HPNSSYGAT
 151 KAAVSLFSSA LKCDLIGLPI RYTELEPGRV ETNLRYTALG KEGAKKRLYD
 201 DNEAIOPHMA ARLTLTALEL PDEVYTRLE VMPTQGVYVG AQMSKLSR

!!AA_SEQUENCE 1.0
 ID YATG_RHISN STANDARD: PRT: 231 AA.
 AC P55661;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE AMINO-ACID ABC TRANSPORTER PERMEASE PROTEIN YATG.
 GN YATG.
 OS Rhizobium sp. (strain NGR234).
 OG Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.

201 VGGIYLITL VASALVRIYD VNLPRGVPL R -

!!AA_SEQUENCE 1.0
 ID Y503_METUA STANDARD: PRT: 406 AA.
 AC 057926;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PUTATIVE 2-ISOPROPYLMALATE/HOMOCITRATE SYNTHASE M0503 (EC 4.1.3.-).
 GN M0503.
 OS Methanococcus jannaschli.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERAVANGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KANE B.P., BORODOVSKY M.,
 RA KLINK H.-P., FRASER C.M., SMITH H.O., MOSESE C.R., VENTER J.C.:
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RA jannaschli".
 RT Science 273:1058-1073(1996).

YATG_RHISN Length: 231 February 14, 2000 08:03 Type: P Check: 6923 ..

1 MLGFTWDG NGELAFALSI LPMILGLIT TLQAFLEGF VACVLGAWFA
 51 VLGRKTRFW AMPRAVLIEF IRDPLLAQL FELYVLPVEY GIIPPAFLTG
 101 ALALGICVSA YISEVYRGI QAVDHGOREA AKSLDPPAR TFWHILPQA
 151 IPRVIALGN YLVSIMKQVP VLSVTIYEM LNAKILIGQ TENYLVPLSM
 201 VGGIYLITL VASALVRIYD VNLPRGVPL R -

!!AA_SEQUENCE 1.0
 ID Y503_METUA STANDARD: PRT: 406 AA.
 AC 057926;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PUTATIVE 2-ISOPROPYLMALATE/HOMOCITRATE SYNTHASE M0503 (EC 4.1.3.-).
 GN M0503.
 OS Methanococcus jannaschli.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERAVANGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KANE B.P., BORODOVSKY M.,
 RA KLINK H.-P., FRASER C.M., SMITH H.O., MOSESE C.R., VENTER J.C.:
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RA jannaschli".
 RT Science 273:1058-1073(1996).

CC -1 SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67500; AAB98494.1; -
CC TIGR: M0703.1; -
CC PROSITE: PS00816; AIPM_HOMOCIT_SYNTH_1; 1.
CC PIR: P00682; HMGL-like; 1.
CC KW Hypothetical protein; Lyase.
CC SEQUENCE 406 AA; 45364 MW; 38573C5F CRC32;
Y503_METUA Length: 406 February 14, 2000 08:03 Type: P Check: 6767 ..
1 MTKVLMEMD FLFENSWKAV CPYNPKLDLK DIVYDITL R DGEQFPGVCF
51 TREOKLEIAR KLDELGLKQI EAGFPVSEK EADIVKTIAN EGINADILAL
101 CRLAKKIDIK AIECDVDGII TFATSPHL KYRPNKSLD EILEMGVEAV
151 EVAKHGLFV AFSADATRT PIEDIKVHK AAEAGADRV HADDTGCAT
201 PQSMFICKI LKENLKKAKHI GVCHNDSEF AVINSIIGLI GSKANSTIV
251 NGIGERAGA ALEELIMALT VLYDVLGIN LEVPELCRM VEYSIKRP
301 KKKPIGELV FAHESGIHVD AVIENPLTE PFLEKRGK RNLILKSHG
351 CRAVAIKKLK MGIDYREML CEIVKVKET REESKFTIDE VKREIVEVL
401 KRRNRN
11AA_SEQUENCE 1.0
ID Y70A_METUA STANDARD; PRT; 102 AA.
AC P83111;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN M0703.1.
GN M0703.1.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales; Methanococcaceae;
OC Methanococcus.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN: JAL-1 / DSM 2661 / ATCC 43067;
CC MEDLINE: 96337999;
CC BUT C. J. WHITE O., OLSEN G. J., ZHOU L., FLEISCHMANN R. D.,
CC SUTTON G. G., BLAKE J. A., FITZGERALD L. M., CLAYTON R. A., GOCAYNE J. D.,
CC KERAVAGE A. R., DOUGHERTY B. A., TOMS J. F., ADAMS M. D., REICH C. I.,
CC SCOTT J. L., GEOGAGAN N. S. M., WEIDMAN J. F., FUHRMAN J. L., NGUYEN D.,
CC UTTERBACK T. R., KELLEY K. M., HURST M. A., PETERSON J. D., SADOW P. W., HANNA M. C.,
CC KLEIN H. P., FRASER C. M., SMITH H. O., MOESE C. R., VENTER J. C.;
CC "Complete genome sequence of the methanogenic archaeon, Methanococcus
CC jannaschii.";
CC Science 273:1058-1073(1996).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67517; AAB98706.1; -
CC DR TIGR: M0703.1; -
CC KW Hypothetical protein; Transmembrane.
CC TRANSMEM 14 34 POTENTIAL.
CC TRANSMEM 35 55 POTENTIAL.
CC TRANSMEM 76 96 POTENTIAL.
CC DOMAIN 76 89 ILE-RICH.
CC SEQUENCE 102 AA; 11685 MW; 2336A893 CRC32;
Y70A_METUA Length: 102 February 14, 2000 08:03 Type: P Check: 4119 ..
1 MGNMNRDK IKSINMWNF IKRITITVGI VISAFAFTIS ILMGMLFIL
51 FLITITFSKT IRLIKSKER SYGLILSLIT GSIITISIV YSHCYIEFKL
101 LI
11AA_SEQUENCE 1.0
ID Y788_BORBU STANDARD; PRT; 440 AA.
AC O51728;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL PROTEIN B90788.
GN B90788.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria: Spirochaetales; Spirochaetaceae; Borrelia.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN: ATCC 35210 / B31;
CC MEDLINE: 98065643;
CC FRASER C. M., CASDINS S., HUANG W. M., SUTTON G. G., CLAYTON R. A.,
CC LATHIGER R., WHITE O., KETCHUM K. A., DOBSON R., HICKET E. K., GWINN M.,
CC DOUGHERTY B., TOMS J. F., FLEISCHMANN R. D., RICHARDSON D.,
CC PETERSON J., KERAVAGE A. R., QUACKENBUSH J., SALZBERG S., HANSON M.,
CC VAN VUUT R., PALMER N., ADAMS M. D., GOCAYNE J. D., WEIDMAN J.,
CC UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
CC GARLAND S., FUJII C., COTTON M. D., HORST K., ROBERTS K., HATCH B.,
CC SMITH H. O., VENTER J. C.;
CC "Genomic sequence of a Lyme disease spirochete, Borrelia
CC burgdorferi.";
CC Nature 390:580-586(1997).
CC -1 SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AEO01177; AAC67121.1; -
CC DR TIGR: B90788; -
CC KW Hypothetical protein
CC SEQUENCE 440 AA; 51471 MW; D44B6E4A CRC32;
Y788_BORBU Length: 440 February 14, 2000 08:03 Type: P Check: 1619 ..
1 MHELDENIOI KIDKEYKNS LDKNRVIAF SGGADSTALL LNKYILSN
51 VIAFYAHLI RSDNEONOEI EHVKGFCIDY NIALQIKKCD IDIKSESARL
101 GVAIEELARK FYIALENAL KENVANVIAL AHENDOJET IIMRFGOSF
151 LDGSGIPSV NRNIIRPLE VSRLEIENEL SLNNGFVVD STNAOINLYLR
201 NRVNNMLPA IKVYKGYEK CLKRISEFSK EFADYFGDE FPFVEGKYY
251 YSDLTATFDL PPKYIVFRLI FKILNBSGIA AKVSKALNE ARVVEINRKK

301 NNVLATNDP FLEKRNKIN LIFKDEKY KPPDILEVG KWSLSIGKI
 351 LKYLECNMA SVSRILKCSY EFRYKPKDR LKAKFFSKF IRCNPATLMT
 401 LALDNLIGI IDLNTLNTW SEKSILKIN ISLIGGLKE
 11AA-SEQUENCE 1.0
 ID Y988_METUA STANDARD: PRT: 170 AA.
 AC P81233;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL PROTEIN M0795.1.
 GN M0795.1.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1/DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UUTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO M1249.1, M0210.1 AND M0785.1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67524; AAB98801.1; -
 DR TIGR: M0795.1; -
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 6 POTENTIAL.
 FT TRANSMEM 31 POTENTIAL.
 FT TRANSMEM 91 POTENTIAL.
 FT TRANSMEM 111 POTENTIAL.
 FT TRANSMEM 170 AA; 19689 MW; 0BF2A40D CRC32;
 SQ
 Y988_METUA Length: 170 February 14, 2000 08:03 Type: P Check: 2865 ..
 1 MKGINPEFY IGMALILASI VSILITIKSI LFLILAFGS LVGITILILI
 51 ISRIKLKIDK GRLKREVKRI FGNRYVKILR LMLVLGAGF IYFSGTFYNS
 101 AVLEFIFIVA FTISEFYKTY RRIYEKIL IEGIAFYWE ELEKTNMDK
 151 NOTLIKIKGI PKRIIVNEII
 11AA-SEQUENCE 1.0
 ID Y983_MYCTU STANDARD: PRT: 282 AA.
 AC P71537;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1999 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 30.9 KD PROTEIN RV0935C.
 GN RV0935C OR MYCY10D7.21.
 OS Mycobacterium tuberculosis.
 OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV
 RX MEDLINE: 96259997.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TERKLA F.,
 RA BADDOCK K., BASHAM D., BROWN D., CHILLINGWORTH I., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOUL S., MORPH L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAUNDREEM M.A., RODERS J.,
 RA RUTHER S., SEEDER K., SKELTON S., SQUARES S., SQUARES R., SULTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: TO M.TUBERCULOSIS RV2161C AND RV3079C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z79700; CAB01997.1; -
 KW Hypothetical protein
 SQ SEQUENCE 282 AA; 30913 MW; 6DBA1010 CRC32;
 Y983_MYCTU Length: 282 February 14, 2000 08:03 Type: P Check: 883 ..
 1 MHYGLVFTS DRGITPAANA RLAEHSQFRT FYVEPHTHIP VKROAHFPT
 51 GDASLPDDRY MRLDPWVSL GAASAVTSRI RLATAVALPV EHPDITLAKS
 101 IATLDLSHG RVSVGVGFGW NTDELVDHGV PEGRRRTMLR EYLEAMALW
 151 TOEACDYDE FVKEGSPWAM PKFVOPHIV LVGAAGTEKN FKWLARSADG
 201 WITTPDVDI DEPVYLLQDI WAAAGRDGLP QIVALDVPRV PDKLARVAEL
 251 GVTVEVFGMP DRASADDAAY VERLAKIAC CV
 11AA-SEQUENCE 1.0
 ID Y988_METUA STANDARD: PRT: 329 AA.
 AC 058395;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL PROTEIN M0998.
 GN M0998.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UUTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: SOME: TO M.JANNASCHII M11633.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67541; AAB98990.1; -
DR TIGR: M09988; -
DR PIR: P01368; DRH; 1.
KW Hypothetical protein.
SQ SEQUENCE 329 AA; 37316 MW; 8B57DAAE CRC32;

Y988_METUA Length: 329 February 14, 2000 08:03 Type: P Check: 9981 ..

1 MELLEYLRKD EYFLCHHNA DPAVSCYA LKYLALQNP NGFRISADS
51 VSKLSRNILN EIGERVDEI YPKLPETVFI VDTASINOLK VNFDELKERE
101 VLIIDHKKKT DLADICKYII IKEDYPTSE IIAEIFKELN IFFPKVVRIA
151 LKGLIYDPR HLKLANSKTF ELISYLKNDI SFQKLYLLS QESDVSKRTA
201 HLKASRMEI REFDKRIAL SHVSHASAC AKTIVSIAD VAVVAVRRK
251 EKEIRVSARC RKHSKYVHL GNMKEIKGE LGSAGGHSF AGGLNAPYDK
301 SSKSEKVIKE VLNLCYKRF EYKRAKON

11AA_SEQUENCE 1.0
ID Y988_PSEAE STANDARD: PRT: 122 AA.

AC P21483; 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE HYPOTHETICAL 13.0 KD PROTEIN IN ALG83 REGION.
OS Pseudomonas aeruginosa.
CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
CC [1]
RN PSEUDOMONAS
RP SEQUENCE FROM N.A.
RX MEDLINE: 90108714.
RA KATO J., CHU L., KIRANO K., DEVAULT J.D., KIMBARA K.,
RA CHAKRABARTY A.M., MISRA T.K.;
RT "Nucleotide sequence of a regulatory region controlling alginate
RT synthesis in Pseudomonas aeruginosa: characterization of the alg83
RT gene." Gene 84:31-38(1989).
RU -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M30145; -; NOT_ANNOTATED_CDS.
DR PIR: J00150; J00150.
KW Hypothetical protein.
SQ SEQUENCE 122 AA; 12967 MW; FEF98CF0 CRC32;

Y932_PSEAE Length: 122 February 14, 2000 08:03 Type: P Check: 6657 ..

1 MALASDASR CCAASASRMP ATACCPPTGA SRTCARSRIP WIMRWITIS
51 TATVACASR RHMPPKRRT TYLPMRACIP NMTPTPTVPP MHRASGMPD
101 SASASRMSA GSVISPAVCG CA

11AA_SEQUENCE 1.0
ID Y932_PSEAE STANDARD: PRT: 216 AA.
AC P39220; P35632;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 24.8 KD PROTEIN IN DULA-RLOA INTERGENIC REGION.
GN YABP.
OS Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RC MEDLINE: 92334977.
RA YURA T., MORI H., NAGAI H., NAGATA T., ISHIIHAMA A., FUJITA N.,
RA ISONO K., MIZOBUCHI K., NAKATA A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region." Nucleic Acids Res. 20:3305-3308(1992).
RU -----
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G., III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER P.D., RODE C.K., MAYHEW G.F.,
RA MAT B., SHAO Y., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RT "The complete genome sequence of Escherichia coli K-12." Science 277:1233-1238(1997).
RU [3]
RN IDENTIFICATION.
RP RUDD K.E.;
RL Unpublished observations (NOV-1994).
RU -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10483; -; NOT_ANNOTATED_CDS.
DR EMBL: AEO00116; AAC73167.1; -
KW ECOGENE: EG12610; Yabp.
DR Hypothetical protein.
SQ SEQUENCE 216 AA; 24791 MW; 89CF19EB CRC32;

YABP_ECOLI Length: 216 February 14, 2000 08:03 Type: P Check: 98 ..

1 MKVSPGMPV TLNMSKNDI YKWSGDKMD YKKNIFORIM ETLNHLWSD
51 KOTAYVILF NFVNNOTGNI NASEYFTGAI NENEREKFIN SLELFNKLKT
101 CAKNPDELVA KGNMRVAVT FGDIELSVTF FLEKKNICTO TLQJHKGQGN
151 LGVLDLRAYL PGVDMEDCYL GKKTGKNSD IYERPGWNA NLGVLPRTVL
201 PRIVLRITVL TWTYLP

11AA_SEQUENCE 1.0
ID Y932_PSEAE STANDARD: PRT: 192 AA.
AC P26166;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE HYPOTHETICAL 20.5 KD PROTEIN IN BCHF-CRTU INTERGENIC REGION (ORF192).
OS Rhodospirillum rubrum (Rhodospirillum rubrum capsulatus).
CC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
CC Rhodospirillum.
CC [1]
RN SEQUENCE FROM N.A.
RP BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.F.;
RA Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
RU -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 21165; CAA77528.1; -
DR PIR: S17812; S17812.
KM Photosynthesis: Hypothetical protein.
SQ SEQUENCE 192 AA; 20551 MW; F1BDB69B CRC32;

YBCT_RHOCA Length: 192 February 14, 2000 08:03 Type: P Check: 2127 ..

1 MDLFFDEFA AHVPEEMAT HIPEPAQOI GAAMDSDRG FQVYIAISR
51 LQELHAIQT LYTADSVGCA NGAYVLIVP PGEQHILGAL IYAMLERRRG
101 VSVRIYFAPG LSDLSRLMAT TRPDALITV GSDRVEICA KIVKTLSSLT
151 KGRNRVAIGG AIYSQRAEAL ARTGADLVN DLSLVISEFS LV

!!AA_SEQUENCE 1.0
ID YBCT_BACSU STANDARD; PRT: 249 AA.
AC O31458;
DR 15-JUL-1998 (Rel. 36, Created)
DR 15-JUL-1998 (Rel. 36, Last sequence update)
DR 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 27.3 KD PROTEIN IN GLTP-CMLJ INTERGENIC REGION.
NM YBCT.
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA HAGA K., LIU H., YASUMOTO K., TAKAHASHI H., YOSHIKAWA H.;
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
RT Bacillus subtilis chromosome", to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases
-1- SIMILARITY: BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE
CC ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB006424; BAA33133.1; -
DR EMBL: 299105; CAB12030.1; -
DR HSSP: P09375; 1HOT.
DR SUBTILIST: BG12746; YBCT.
DR PROSITE: PS01161; G1C.GALNAC.ISOMERASE. 1.
DR PFAM: PF01182; Glucosamine Iso. 1.
KW Hypothetical protein; Isomerase.
AC STPE
NM YBCT.
OS SEQUENCE 249 AA; 27289 MW; 20883DBC CRC32;

YBFT_BACSU Length: 249 February 14, 2000 08:03 Type: P Check: 3388 ..

1 MKILAEHYE ELCKLSAII KEQIOAKKDA VLGANGSTP VELYQOLISD
51 YQAGEIDFSK VTFENLDEYA GISPSHPOSY NHEMHEHLFO HINMOPDHIH
101 IPOGDNPOLE AACRYEDLI RQAGSIDVOI LGISANGHIG FNEPSPDEED
151 RTYRVKLSIES TIOANAFETG GPPVLYPRLA ISMGIKTIME PSKHIVLLAS
201 GEEKADAIOK MAGPYTIDV PASILOKHNH VTIADYKAA OKLSASFS

!!AA_SEQUENCE 1.0
ID YC08_KLEPN STANDARD; PRT: 373 AA.
AC Q48454;
DR 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DR 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 42.6 KD PROTEIN IN CPS REGION (ORF8).
OS Klebsiella pneumoniae.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHEID;
RX MEDLINE: 95204345.
RA ARAKAWA Y., MACHAROTAYANKUN R., NAGATSUKA T., ITO H., KATO N.,
RA OHTA M.;
RT "Genomic organization of the Klebsiella pneumoniae cps region
RT responsible for serotype K2 capsular polysaccharide synthesis in the
RT virulent strain Cheid.",
RL J. Bacteriol. 177:1788-1796(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D21242; BAA04779.1; -
DR PFAM: PF00534; Glycos_transf_1; 1.
KM Hypothetical protein.
SQ SEQUENCE 373 AA; 42641 MW; 4150C919 CRC32;

YC08_KLEPN Length: 373 February 14, 2000 08:03 Type: P Check: 4280 ..

1 MNILVNTLX YPKKIGAEV SVQLAESLI EKHSTVVS IHENNEKDI
51 EHNQVKIYL PPSNIYNGLS LTRKNPLSKI LMHLIDLYNF KIAEFENII
101 MDVAPDIEHT NNLGSGICAV WOKAKYKCR VHTSHDIYL IHPCKLKYKN
151 GSENSYKSLA VSLMSLSKRI LKKNVDYVG ISNYIKDKHI EAGFKSTK
201 YTIYNSYKSN VILDTIAND KRLGFIQRLT YKGGPDQFC LAQINKTKKF
251 IAGEEPDKNS ASLKQALDS NVELGYPV DDFMOKVDII VLPKMOEPP
301 GRVVEAIFA GRVVLTRNGV GTELSRILP NIYFLEDIOD IDSIPFVEI
351 DDESKRIFNV DYTEQYLKI YKG

!!AA_SEQUENCE 1.0
ID YC21_METJA STANDARD; PRT: 299 AA.
AC O58618;
DR 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DR 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1221.
GN MJ1221.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96537999.
RA BOLT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.E., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SAOOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KATNE B.P., BOROVOSKY M.,

```

RA KIENK H.-P., FRASER C.M., SMITH H.O., MOESER C.R., VENTER J.C.:
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RL jannaschii, 1996." J. BACTERIOL. 178:1058-1073(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO M.JANNASCHII M0678.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U67563; AAB9234.1; -
 DR TIGR: M1221; -
 KW Hypothetical protein: Transmembrane.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 261 281 POTENTIAL.
 SQ SEQUENCE 299 AA; 32538 MW; 8E8A08F7 CRC32;
 YC21.METRA Length: 299 February 14, 2000 08:03 Type: P Check: 2920 ..
 1 MMEGERMRYM KIIPKFLN TVEIILKNN AYSIIIEPL KTSIEDGIII
 51 TGNADARAE KIVLEKLKLG LSEKGGSVT IMPANITSC REEGIASTSL
 101 SPLEIYKAK TMVKITKNVI IKVILASMG VIGLIEHNIP TLIGAMITAP
 151 LVDTWGSNI GTVLGKELF IQCKKELLC SGIVIVCAFI PSELEFSKEL
 201 VLYVISETSI ILSAIVAIIA GISGSGMSIAS GKEVEIIISVT IDVSILIPAL
 251 LMGMALATMD LYLIVTEPIL LAIVIVLDV GGVIGLKRYV GKINOKIKY
 !!AA_SEQUENCE 1.0
 ID YC38.PORPU STANDARD; PRT; 291 AA.
 AC P51321;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEICAL 32.4 KD PROTEIN YCF38 (ORF291).
 GN YCF38.
 OS Porphyra purpurea.
 OC Chloroplast.
 OG Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AVONPORT;
 RA REITH M.E., MONHOLLAND J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 RL genome.";
 CC -1- SIMILARITY: BELONGS TO THE YCF38 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U38804; AAC08207.1; -
 DR MENDEL: 10386; PORPU:YC38.1;
 KW Chloroplast; Hypothetical protein.
 SQ SEQUENCE 291 AA; 32427 MW; 921437DB CRC32;
 YC38.PORPU Length: 291 February 14, 2000 08:03 Type: P Check: 4681 ..

1 MTFASKKIE LKPIKFTP KYVSEYIHO IEALVORLEL QVWRPAILM
 51 AGIIQPLML VLEGGLEFCA PUNLEIINTS YNFFLSGGII VETSTGALN
 101 SGLPLMDRE FGFLNLLLA PLISRTIIE SSATFMTICS LIQVIFYIA
 151 SLFQNSLS SNTLIFALI VLIVYGYTM LSLALFTLP GHIELLAIL
 201 VYNPLEFSS TALAPIYFAP PWQLIASLN PLSTALEGIR YIYSTDMNF
 251 TESVIRKSWG DISLGOITSL LIPLDYIGAY IVSNILKARL N
 !!AA_SEQUENCE 1.0
 ID YCBO.ECOLI STANDARD; PRT; 182 AA.
 AC P75855;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL FIBRIAL-LIKE PROTEIN IN PEBN-PYRD INTERGENIC REGION
 DE PRECURSOR..
 GN YCBO.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12/MG1655;
 RA MEDLINE: 97426617;
 RA BLATTNER F.R., FLOURENT G., III, BLOCH C.A., PERA N.T., BURLAND V.,
 RA RIESE M., COLLADO-VIDES J., GLASNER F.D., ROSE D.J., MAHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000196; AAC74024.1; -
 DR ECOGENE: EG13709; YCBO.
 DR PFAM: PF00419; Fimbrin; 1.
 KW Hypothetical protein; Fimbria; Signal.
 FT SIGNAL 1 182 POTENTIAL.
 FT CHAIN 25 182 HYPOTHEICAL FIBRIAL-LIKE PROTEIN YCBO.
 SQ SEQUENCE 182 AA; 19025 MW; CE190E78 CRC32;
 YCBO.ECOLI Length: 182 February 14, 2000 08:03 Type: P Check: 3224 ..
 1 MITMKKSVLT AFTIVCAIS SYMAADNVAI TDGSVTFNGK VIAPACTIVA
 51 AKRDSVTLF DVSAIRKLQIN GOVSGVQIDV PIEIKDCDT VTKNATFTFN
 101 GTADTQITTA FANQASDDA TNVALOMYNN DGTATITPT ETGNILLDGS
 151 DQITFEKVDY IATGKATSGN VNAVTFNIN YV
 !!AA_SEQUENCE 1.0
 ID YCG2.ECOLI STANDARD; PRT; 78 AA.
 AC P75991;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL 8.8 KD PROTEIN IN ICDC-MINE INTERGENIC REGION.
 GN YCG2.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;


```

51 YKSNMPEVO MENISAFINY AOOVHVPSQ DMFOTSDUFE RRNDQCVLRS
101 IHSFRYAK MFGKVRGLG PKLAERKPRV FSAOQOEFRR EGVNSLOYGS
151 FDMPTGTEK IAFSRDRDPT GNMV

!!AA_SEQUENCE 1.0
ID YDNG_SCHPO STANDARD: PRT: 344 AA.
AC 014220:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 59.8 KD PROTEIN CGB12.16 IN CHROMOSOME 1.
GN SPAC6B12.16.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archizosomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972:
RA GENTLES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 298531; CAB11075.1;
KM Hypothetical protein
SQ SEQUENCE 344 AA; 39810 MW; BBE84116 CRC32;

YDNG_SCHPO Length: 344 February 14, 2000 08:03 Type: P Check: 3748 ..

1 MNSNEIENFI YLNTLTISEV PEPNSVFYTP KCFNSKLPH QSDIEPSSA
51 CSLSKRTIID GGADEPYDSS DSCATEFAD FEKNDLDDFA LDGEIINIH
101 VNQTSPTYTH NAEPDLQIT FSSNLNNSN RRQITCNFM ANDASKENET
151 PYVALNKNFN PVLTEYTOQ HLYQCKMYLE NHITSRPHF YTKLPDVSIV
201 PNMHPYDDE VTAKAPKDD FYVPRFTRGH GISKGLCP1 CSHQGEFIWL
251 RTTSAYWTH MNEVGHSHK GRPYQPIEF RTVRLKTRN AIGVPMKKM
301 IEKCHQCNK WIRQGRKDV SVKIPPEFW RHAHCHIT TDLR

!!AA_SEQUENCE 1.0
ID IEK4_LEAST STANDARD: PRT: 128 AA.
AC 040057:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 14.3 KD PROTEIN IN PTP3-1LV1 INTERGENIC REGION.
GN YER084W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972:
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTWELL G., HUNICKER-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PATEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SMOGGEN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;

```

```

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18839; AAB64639.1;
KM Hypothetical protein.
SQ SEQUENCE 128 AA; 14313 MW; 564479CB CRC32;

YER4_YEAST Length: 128 February 14, 2000 08:03 Type: P Check: 5798 ..

1 MELICITPY HSNLPEFLF FCPSPKRRR GHRKFLTLG YKSNLIRKL
51 LPPSLFTKRV MNPSSHPS PDFPGSSAS PRVKLRPSTL WAPPLTVSSD
101 PAASSSTAP VVTDKPVP AVSKRYQP

!!AA_SEQUENCE 1.0
ID YF8_MYCTU STANDARD: PRT: 148 AA.
AC 010772:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 16.4 KD PROTEIN RV1558.
GN RV1558 OR MYC48.07C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV.
RA MEDLINE: 9825987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGMEIER K., GAS S., BART C.E., ILL, TERKHA F.,
RA BADOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORSNER T., JAGELS K., KROGH A., MCLEAN J., MOULS S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV1261C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 274020; CA98933.1;
KM Hypothetical protein.
SQ SEQUENCE 148 AA; 16347 MW; 7F320812 CRC32;

YF8_MYCTU Length: 148 February 14, 2000 08:03 Type: P Check: 497 ..

1 MPELGEYAPS PLDMREQAD TYKSGGTEG TQLOGKPVIL LTVGAKTK
51 LKRTPLMRE HDGOYAIVAS LGAPRPYV YHNVYKNRPV ELQDGTVTGD
101 YDAREVFGDE KAIMORAVA VMPDYASYOT KTRQIPEVY LIPVRAGG

!!AA_SEQUENCE 1.0
ID YG29_BPSPI STANDARD: PRT: 148 AA.
AC P31653:

```

01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE HYPOTHEORETICAL 16.2 KD PROTEIN IN GENE 29.5 REGION.
OS Bacteriophage SP01
CC Viruses; dsDNA viruses, no RNA stage: Tailed phages: Myoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:WMT
RX MEDLINE: 92351562
RA WILHELM K., ROESER W.;
RT "Deoxyuridylylate-hydroxymethylase of bacteriophage SP01";
RL Virology 189:640-646(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60728; CAA43135.1; -
DR PIR: S21504; S21504.
KW Hypothetical protein.
SQ SEQUENCE 148 AA; 16183 MW; 7C6906B5 CRC32;
YG29_BPSP1 Length: 148 February 14, 2000 08:03 Type: P Check: 7782 ..
1 MNQGVNLKIG GYVVDPPVK VESKTYGELE VTRKKVYHRM LDIIIPVGLA
51 TKTLALSPG ALAAGVDSA DKIRGFHDI IDVTALAEF ILMFVLTMC
101 VLIATKRNA GWERLKNVGT AVAGIALLEPT FFSFLRWVSS IYSSSITF
11AA_SEQUENCE 1.0
ID YGK9_YEAST STANDARD; PRT; 107 AA.
AC P53138;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEORETICAL 12.4 KD PROTEIN IN TAF60-G4P1 INTERGENIC REGION.
GN YGL109W OR G3065.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97197974.
RA PAOLUZI S., MINENKOVA O., CASTAGNOLI L.;
RT "The genes encoding the transcription factor YTAFL160, the G4P1
RT protein and a putative glucose transporter are contained in a 12.3 kb
RT DNA fragment on the left arm of Saccharomyces cerevisiae chromosome
RT YII.";
RL Yeast 13:85-91(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X97644; CAA66243.1; -
DR EMBL: 272631; CAA96815.1; -
KW Hypothetical protein.
SQ SEQUENCE 107 AA; 12435 MW; 349DFDE6 CRC32;
YGK9_YEAST Length: 107 February 14, 2000 08:03 Type: P Check: 6897 ..
1 MAAGNPGLADI QVYKRYKAR RMEGQKNSC TIAITDSLQY YCRRLSHKS

51 CFPSPQHAF SBNPLPSESY EFWHALELAF CLTRPYCTFH SLEISISQOL
101 TLRPPLG
11AA_SEQUENCE 1.0
ID YGIR_STRCO STANDARD; PRT; 66 AA.
AC P05954;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEORETICAL PROTEIN IN GLNR 3 REGION (FRAGMENT).
OS Streptomyces coelicolor.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:A3(2);
RX MEDLINE: 93345814.
RA WRAY L.V. JR., FISHER S.H.;
RT "The Streptomyces coelicolor glr gene encodes a protein similar to
RT other bacterial response regulators.";
RL Gene 130:145-150(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02213; AAA02839.1; -
DR PIR: P06044; P06044.
DR HSSP: P15035; LPRV.
KW Hypothetical protein.
FT NON_TER 66
SQ SEQUENCE 66 AA; 7094 MW; 502F47A4 CRC32;
YGLR_STRCO Length: 66 February 14, 2000 08:03 Type: P Check: 9954 ..
1 MAYTRDDVA RLAGSTAVV SYVINGNRP VAPATREHVL AAIKEIGYRP
51 DRVAQAMASR RDLIG
11AA_SEQUENCE 1.0
ID YHAL_CRYPA STANDARD; PRT; 319 AA.
AC P10941;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHEORETICAL PROTEIN 1 IN HYPOVIRULENCE-ASSOCIATED DS-RNA GENETIC
DE ELEMENT (CONTAINS: P29 PROTEINASE).
OS Cyphonectria parasitica (Chesnut blight fungus) (Endothia
OS parasitica).
CC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
CC Diaporthales; Valsaceae; Cyphonectria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EP713;
RX MEDLINE: 89251594.
RA RAE B.P., HILLMAN B.L., TARTAGLIA J., NUSS D.L.;
RT "Characterization of double-stranded RNA genetic elements associated
RT with biological control of chestnut blight: organization of terminal
RT domains and identification of gene products.";
RL EMO J. 8:657-663(1989).
CC -----
CC MISCELLANEOUS: DOUBLE-STRANDED RNA GENETIC ELEMENTS ARE ASSOCIATED
CC WITH BIOLOGICAL CONTROL OF THE FUNGAL DISEASE CHESTNUT BLIGHT.
CC THIS DS-RNA ARE ASSOCIATED WITH HYPOVIRULENCE. THEY ARE LOCALIZED
CC IN THE CYTOPLASM
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X14524; CA33666.1; -
DR PIR: S03833; S03833.
KW Hypothetical protein; Hydrolase; Thiol protease.
SQ SEQUENCE 319 AA; 35443 MW; 6309763E CRC32;

YHAI_CRYPA Length: 319 February 14, 2000 08:03 Type: P Check: 3269 ..

1 MAOLIKPSQS LVLSSEVDPT TVDPFVSVRT EEVVPAGCIT IMEYDSCGD
51 VPGFLSHGDL RLRLPDPGVC KCQVHFLPT VLKSGSTGTV PEHPAVLAF
101 IGRPRRCSLE QRTKELDSRF LQVHGGLPA RPSYMLARP REVRGLCSSR
151 NSGLAOFGOG YCYLSAIVDS ARWRVARTG WCVRVADYLR LLQWGRRSF
201 GSFOIERSAV DHYHYVYVDA EYQSEODGAL FYQALIGLAE KDPLARIGR
251 LNPILAEPAP GSALRVEPYT PQTERRKGST RMTGRDPTIV FRMGWYGH
301 QHPCSCGYG VEERGRPS

!!AA_SEQUENCE 1.0 STANDARD; PRT; 302 AA.
ID YHCT_BACSU

AC P54604; -
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL 33.7 KD PROTEIN IN CSPB-GLFP INTERGENIC REGION.

GN YHCT
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.

RN
RP SEQUENCE FROM N.A.
RX STRAIN-168;
RX MEDLINE; 97124185.

RA NOBACK M.A., TERPSTRA P., HOLSAPPEL S., VENEMA G., BRON S.;
RT "A 22 kb DNA sequence in the CSPB-91PFXD region at 75 degrees on the
RL Microbiology 142:3021-3026(1996)."

CC -1- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X96983; CA65704.1; -
DR EMBL: X99108; CAB12749.1; -
DR SUBMILIST; B611598; YHCT.
DR PROSITE; PS01129; PSI_RLU; 1.
DR PFM; PF00849; YABO; 1.
KW Hypothetical protein.
SQ SEQUENCE 302 AA; 33740 MW; 2EB71306 CRC32;

YHCT_BACSU Length: 302 February 14, 2000 08:03 Type: P Check: 1929 ..

1 MNOKRGLEI LINEKODGW LPSVLTKALK ASKEVIDMM SHQOIVNHE
51 SVLNMMIVRK GDRVFDLQE SEASSVPEY GELDILFEDN HMLINKPAG
101 IATHNEDGO TGLANLIV HYOINGETCK VRHYRLDOD TSGAIVFAH
151 RLAAHILDOO LEKTLKRTY TAAEGKRT KGTINPPIG RDRSHPTRR

201 VSPGOTAVT HFKWASNAK ERLSVELEL ETGRTHQIRV HLASLGHPLT
251 GDSLYGGGSK LNRQALHAN KQAVHPITD ELIVAEPFP ADMKNLCRTY
301 FS

!!AA_SEQUENCE 1.0 STANDARD; PRT; 429 AA.
ID YHG3_YEAST

AC P38756;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHEICAL 48.9 KD PROTEIN IN RPL14B-6PAI INTERGENIC REGION.

GN YH003C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.

RN
RP SEQUENCE FROM N.A.
RX STRAIN-S288C / AB972;
RX MEDLINE; 94378003.

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATUNG S., GEISEL C., KIRSTEN J.,
RA KUCABA T., HILIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA VAUDIN M.;

RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT viii";
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO YEAST YKL027M, SOME, TO E. COLI MOLYBDENUM
CC COFACTOR BIOSYNTHESIS PROTEIN A (MOA).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U10555; AAB68430.1; -
DR PIR; S46801; S46801.
DR PFM; PF00899; ThLF_family; 1.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 48883 MW; 709A28CF CRC32;

YHG3_YEAST Length: 429 February 14, 2000 08:03 Type: P Check: 5569 ..

1 MANTWKLIA TTAISVFT QLAQVWKEY KLSKANKRK TVSPROYD
51 HLFREGQARN YAFLEGGR KIRQYIVIV GAGEVGSWC TMLIRSGCOK
101 IMIDPENIS IDSLNTHCA VLSDIGRPV QCLKEHLSKI APWEIKARA
151 KATKENSND LIFADGESP FIVDCLDNL SKVDLEVAH HNKIDVISM
201 GVATKSDPTR VSINDSMTE FDPISRCVR KLRKRIAGT ISVFSNEML
251 DPRDDILSP IDENHAINA VRDALRHP ELDTMGING LSIATWILTK
301 VSGTPKENE VKNRLKPYDS ILETFOKMA RLNENKERS LLGLEEYVI
351 VEENFRGKSP ISGYSTKLAL TKWEANKKS LINVVLMTKE EOEIHKAIL
401 LDGEKLTAVY SEEVLDIER LFKEEYIS

!!AA_SEQUENCE 1.0 STANDARD; PRT; 197 AA.
ID YHGN_ECOLI
AC P46851;
DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEICAL 21.5 KD PROTEIN IN ASD-GNTU INTERGENIC REGION (0197).
 GN YHGN.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97428617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE UPF0056 (MARC) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U18997; AA58323.1;
 DR EMBL: AF000420; AAC76459.1;
 DR EMBL: EG12841; YHGN.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 139 159 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 SQ SEQUENCE 197 AA; 21490 MW; 089FBACD CRC32;
 YHGN.ECOLI Length: 197 February 14, 2000 08:03 Type: P Check: 1678 ..
 1 MNEIISAVAL LILIMPLGN LPIFMSVLH TEPKRRRAIM VRELLIALIV
 51 MLVFLFAGER ILAFSLRAE TVSISGILL FLAIKMIFF SASGNSGSLP
 101 AGEFFIYVL AIPVAGPTI LATIMLSHQ YPNQGHLYI ALLANGTIF
 151 VILQSLFL RLIGKGVNA LERLMGLIV MMATOMFLDG IRMMWKG
 11AA_SEQUENCE 1.0 STANDARD; PRT; 164 AA.
 ID Y129_MYCTU
 AC 050604.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE HYPOTHEICAL 18.1 KD PROTEIN RV1829.
 GN RV1829 OR MTC1A11.14C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Filumetes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 98295987.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULÉ S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
 RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z78020; CAB01481.1;
 DR EMBL: Z78020; CAB01481.1;
 KW Hypothetical protein.
 SQ SEQUENCE 164 AA; 18114 MW; 09A1BF7B CRC32;
 Y129_MYCTU Length: 164 February 14, 2000 08:03 Type: P Check: 7074 ..
 1 MGEVAVGIR VEQPQNPVL LLREANGDRY LPIWIGQSEA AAILQOQV
 51 EPPRLTHDL IRDLIALGH SLKEVRIYDL QEGTFYADLI FDRNIVSAR
 101 PSDSVATLR VGVPIVEEA VLAQAGLIP DESDEATRA VREDEVERFX
 151 EFLDSVSPD FKAT
 11AA_SEQUENCE 1.0 STANDARD; PRT; 317 AA.
 ID Y129_MYCTU
 AC 010863.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE HYPOTHEICAL 33.9 KD PROTEIN RV1936.
 GN RV1936 OR MTC139.23C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Filumetes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 98295987.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULÉ S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
 RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: TO B.SUBTILIS XXIF.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z74025; CA98390.1;
 DR EMBL: Z74025; CA98390.1;
 DR PFAM: PF00582; USP; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 317 AA; 33879 MW; AAFB579 CRC32;
 Y129_MYCTU Length: 317 February 14, 2000 08:03 Type: P Check: 1836 ..
 1 MSAQQTNLGI VVGVDGSPCS HTAVENARD AQMRNVALRV VQVVPVITA
 51 PEGNAFYSR FQDAKREIV EHSIVNAQH QIVEQAHKVA LEASSSRRRA

```

101 QITGEVLHGQ IYPTLANISR QVAMVVLGYR GOGAVAGALL GSVSSLVHR
151 AHGPVAVIPE EPRPARPRA PYVVGICGSP TSGLAELIIF DEARRRVDL
201 VALHANSDMG PLDFPRLNKA PIEKRNLEDE QEKMLARRLS GMQRIEDYV
251 VHKVYVCDRP APRLELAOT AQLVVGSGH RGGFGPMHIG SVSRVAVNSG
301 QAPYIVARIP QDPAPVA

11AA_SEQUENCE 1.0
ID YJXJ_ECOLI STANDARD: PRT: 173 AA.
AC P39410:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 18.6 KD PROTEIN IN TPRP-GPMB INTERGENIC REGION (F173).
GN YJXJ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 95334362
RA BURLAND, J.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATINER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RT Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE OF 31-173 FROM N.A.
RX MEDLINE: 81053831.
RA SINGLETON C.K., RODDER W.D., BOGOSIAN G., SOMERVILLE R.L., WEITH H.L.;
RT "DNA sequence of the E. coli trp gene and prediction of the amino
RT acid sequence of the Trp repressor."
RT Nucleic Acids Res. 8:1551-1560(1980).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; AAA97290.1;
DR EMBL: AE000509; AAC7947.1;
DR EMBL: J01715; NOT_ANNOTATED_CDS.
DR ECGENE: EG12600; yjxj.
KW Hypothetical protein.
SO SEQUENCE 173 AA; 18570 MW; 2DC3DAB3 CRC32;

YJXJ_ECOLI Length: 173 February 14, 2000 08:03 Type: P Check: 3873
1 MLINHGVCA TTNPAKIOAI LQAFHEIRGE GSCHIASVAN ESGVEQDFG
51 SEETRAGARN RVANARRLP EADFWAIEA GIDGDSFVW VVIENASORG
101 EARSATPLP AVILEKVRG EALGPVMSRY TGIDEIGRKE GAIGVFTAGK
151 LTRASVYHQA VILALSPFN AVY

11AA_SEQUENCE 1.0
ID YJXJ_ENTAE STANDARD: PRT: 54 AA.
AC P39430:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHETICAL PROTEIN IN TPRP 3 REGION (FRAGMENT).
GN YJXJ.
OS Enterobacter aerogenes (aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Enterobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94268903.
RA ARVIDSON D.N., ARVIDSON C.G., LAMSON C.L., MINER J., ADAMS C.,
RA YOUNDERMAN P.;
RT "The tyrophan repressor sequence is highly conserved among the
RT Enterobacteriaceae."
RT Nucleic Acids Res. 22:1821-1829(1994).
RN [2]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L26582; AAC36893.1;
KW Hypothetical protein.
KW NON_TER
FT SEQUENCE 54 AA; 5794 MW; 0E5A3E2A CRC32;

YJXJ_ENTAE Length: 54 February 14, 2000 08:03 Type: P Check: 4167
1 EALGPVMSGH TGIDQIRKE GAIGVFTAGK LTRSVYHQA VILALSPFN
51 AIYR

11AA_SEQUENCE 1.0
ID YJX8_YEAST STANDARD: PRT: 196 AA.
AC P46589:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 21.6 KD PROTEIN IN ATP12-PL17B INTERGENIC REGION.
GN YJL17BC OR J0490.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA OBERHAIRER B., PIRAVANDI E., RINKER M., DOMDEY H.;
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49453; CAA89473.1;
KW Hypothetical protein; Transmembrane.
KW TRANSMEM 125 145 POTENTIAL.
FT SEQUENCE 196 AA; 21605 MW; AEA06F63 CRC32;

YJX8_YEAST Length: 196 February 14, 2000 08:03 Type: P Check: 8234
1 MLCGLTIVIL PGKDAITQI IDFDKNIGFN VEETESALTL TLKGATWGAN
51 SFDKLEFQC NDNMKODELT SHTWADKSIO LTLKGPSGCL KSKDDDKNG
101 DGDNGKDGDS EGKPKAKAG GTSMFTWLF YALFTLLYL MYVSFLNTRG
151 GSFQDFRAEF IQRSQFLTS LPEFCREVS RILGRSTQR GGVSAV

11AA_SEQUENCE 1.0
ID YJXFP_ECOLI STANDARD: PRT: 79 AA.
AC P35677:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

```

15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 9.0 KD PROTEIN IN PROA-PERR INTERGENIC REGION.
YKFF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617;
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GIANER F.D., RODE C.N., MAHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOLDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.,
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000133; AAC73352.1;
DR ECOGENE: EG14283; YKFF.
KM Hypothetical protein.
KW SEQUENCE 79 AA; 9014 MW; F01375F7 CRC32;
YKFF_ECOLI Length: 79 February 14, 2000 08:03 Type: P Check: 8141 ..
1 MTOSVLLPG PFTROQAV TTYTSNTLE DDGSHFRVY VRDREGMW
51 RANNEPDAG EGLNRYIRS GIRDTMTR
11AA SEQUENCE 1.0
ID YN27 MARPO STANDARD: PRT; 69 AA.
AC P38469;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE HYPOTHETICAL 7.9 KD PROTEIN IN NAD6-NAD3 INTERGENIC REGION (ORF 69).
GN YN27.
OS Marchantia polymorpha (Liverwort).
OC Marchantiophyta.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;
OC Marchantiales; Marchantiaceae; Marchantia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92114051.
RA ODA K., YAMATO K., OHTA E., NAKAMURA Y., TAKEKURA M., NOZATO N.,
RA AKASHI K., KANGAE T., OGURA Y., KOHCHI T., OHYAMA K.,
RT "Gene organization deduced from the complete sequence of liverwort
RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT mitochondrial genome."
RL J. Mol. Biol. 223:1-7(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M68929; AAC09407.1;
DR PIR: S25969; S25969.
DR MENDEL: 2099; MARPO: YN27.1.
KW Mitochondrion; Hypothetical protein.
SQ SEQUENCE 69 AA; 7940 MW; 699EE385 CRC32;
YN27_MARPO Length: 69 February 14, 2000 08:03 Type: P Check: 7679 ..

1 MASLSTPRT QTLVPAHTF IYPAHTSFH GPAFGQALH LFGPLGLFE
51 RENACHIFD NRKTPRLQ
11AA SEQUENCE 1.0
ID YN06 YEAST STANDARD: PRT; 139 AA.
AC P53842;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 15.5 KD PROTEIN IN P1K1-POL2 INTERGENIC REGION.
GN YN266W OR N0800.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE: 96310631.
RA SEN-GUPTA M., LYCK R., FLEIG U., NIEDENTHAL R.K., HEGEMANN J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes."
RL Yeast 12:505-514(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X92494; CA63232.1;
DR EMBL: Z71542; CA96173.1;
KM Hypothetical protein; Transmembrane.
FT DOMAIN 25 33 POLY-SER.
FT TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
SQ SEQUENCE 139 AA; 15510 MW; 11A8B5FA CRC32;
YN06_YEAST Length: 139 February 14, 2000 08:03 Type: P Check: 7226 ..
1 MMLNHTYKL LSYFLRKASN RFENSSSSF SCSEFLVFLV VFEDCFST
51 TSFLISFGIL SSFLIFSLFC LGFLTVIGCL ASALSLSIS KAKIGFSSSL
101 SSISPESSLK SEEMLEDD KESSLLYGT STVPAISRK
11AA SEQUENCE 1.0
ID YN11 CAEEL STANDARD: PRT; 170 AA.
AC P34496;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 19.1 KD PROTEIN PAR2.1 IN CHROMOSOME III.
GN K02D10.5 OR PAR2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WILSON R.;
RT Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE OF 53-170 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718;
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COOPER T., COOPER J., COLLISON A.,
RA CAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,

RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSON L., JONES M., KERSHAM J., KIRSTEN J., LAISTER N.,
 RA LARPERLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SUTHERLAND J., TIERNEY M., SMITH A., SMITH M., SONNHAMMER E., SPADEN R.,
 RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLMAN P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans 368332-38(1994).
 RL Nucleotide 368332-38(1994).
 CC -1- SIMILARITY: TO PLASMID RK2 P116.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D00025; AAA50616.1;
 DR EMBL: I14710; AAA28082.1; ALT_INIT.
 DR PIR: S44839; S44839.
 DR WORMPEP: PAR2.1; CE00768.
 DR PRAM: PF00436; SSB: 1.
 KW Hypothetical protein.
 SO SEQUENCE 170 AA; 19182 MW; 6C09F37E CRC32;
 YNL1-CAEEL Length: 170 February 14, 2000 08:03 Type: P Check: 5825 ..
 1 MMSLSTISK STVRCMSLTS KMAEDPSKQ EYDDJFAKRP QHNHPEDRRH
 51 AYSNNEVELY GGVALDPLKX TGRNGRPYLI FNILINSYK QQDSTLIDQT
 101 ERRAVSVFQK QAEILSKTIK KGSRLWQGR LHYSGGQNDQ QGNNTQNTY
 151 IINQTVQPLA RAARENPDQH
 11AA_SEQUENCE 1.0 STANDARD; PRT: 336 AA.
 ID YNC9 CAEEL
 AC P34542;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE HYPOTHEICAL 37.5 KD PROTEIN R05D3.9 IN CHROMOSOME III.
 GN R05D3.9
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 PN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
 RA CAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSON L., JONES M., KERSHAM J., KIRSTEN J., LAISTER N.,
 RA LARPERLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SUTHERLAND J., TIERNEY M., SMITH A., SMITH M., SONNHAMMER E., SPADEN R.,
 RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLMAN P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans 368332-38(1994).
 RL Nucleotide 368332-38(1994).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07144; AAA28134.1;
 DR PIR: S44865; S44865.
 DR WORMPEP: R05D3.9; CE00288.
 KW Hypothetical protein.
 SO SEQUENCE 336 AA; 37480 MW; 41DDA58 CRC32;
 YNC9-CAEEL Length: 336 February 14, 2000 08:03 Type: P Check: 399 ..
 1 MTLIFVRAS TLKTIILSR IEMGRILEKP RILGTWSFTK IGLNCESLSG
 51 AIGRLPNESK LSEFKKGIYG GKRRSVYIG VFPSDQDSF QHFOIAEKL
 101 SGKYYVAFI NENSKPAITT YRHAEKOKRT DYSGKRDPAI LMEFITKSSI
 151 PSIIIGISGF TIDLPLEHOKR PILIQIGAG EPEPEFKLSA RODARKTYIF
 201 TKIGESSEMI KTKRKALGVG ENESIVFELN KDEVHRIPIIS KPKCPDHQK
 251 ILQMIITSEA DQVLSKKEPH PLRYLYVEKV NEVFGFEETL VLPDHTLELD
 301 SDPFSRHPII TEGGGGGCPF MGGGAPTES SQHSEL
 11AA_SEQUENCE 1.0 STANDARD; PRT: 381 AA.
 ID YN06 CAEEL
 AC P50444;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 42.9 KD PROTEIN R74.6 IN CHROMOSOME III.
 GN R74.6
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 PN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX GARDINER A.;
 RA Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: TO DROSOPHILA PELOTA, YEAST DOM34 AND M.JANNASCHII
 MO0174.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z36238; CA85227.1;
 DR WORMPEP: R74.6; CE01059.
 DR PFAM: PF01605; RFL: 1.
 KW Hypothetical protein; Cell division; Meiosis; Mitosis;
 KW Nuclear protein.
 SO SEQUENCE 381 AA; 42878 MW; 34540990 CRC32;
 YN06-CAEEL Length: 381 February 14, 2000 08:03 Type: P Check: 431 ..
 1 MKQFARGIER DGTGFVLA EEAEDMHIY NLIRIGDIIR ASTIRKVAE
 51 TSTGTTSSOR VHTMLTVSVE SIDFPGQAE LHKGRNIEE NDIVVLGAYH
 101 TIDLEPNRKF TLQTEWDSI DIERLNLAD PAQADVAAY VJHEBLANVC
 151 ITPPMVTLR AKIDMTIPRK RKGFSSQHEK GLEKFEAVS TAFMHNVLQ
 201 VKCVIVASR GFVKDAFMQH LIAHADANGK KFTTEORAKF MLTHSSSGFK

```

251 HALKEVLEP OVALRLADTK AGEVKALNQ FLELMSTEDP RAYGFNHVN
301 RANOELEIET LTVADSLFRA ODIEITRRKY RLVESVREGN GKVIFFSMH
351 VSGEOLAQLT GCAILIRFPM PDLDEPMDE N

!!A-SEQUENCE 1.0
ID YOTC_CAEEL STANDARD: PRT: 232 AA.
AC P19296:
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE HYPOTHETICAL 26.8 KD PROTEIN.
OS Thermoproteus tenax virus 1 (strain KRA1) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC Lipothrixvirus.
RN (1)
RP SEQUENCE FROM N.A.
RA NEUMANN H.;
RL Submitted (MAR-1989) to the EMBL/Genbank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X14855; CAA32992.1;
DR Hypothetical protein.
KW SEQUENCE 232 AA; 26862 MW; E457433C CRC32;
SQ

YOTC_TTV1 Length: 232 February 14, 2000 08:03 Type: P Check: 9154

1 MDVAVALYS RDDALDLRL TLHYWKRSD GLRGQIYVS NTNVLLPRRA
51 KLIOQRKQV GGARCDALIES TSEDYIIFSD GHVTPPREIS KMLVEPWSV
101 PINHILPFG TSGRAYSLP FMPDERQFLM CSCSKYKEEI TTGEPIYAM
151 SKRYSMSKC YTSYGLDLFQ FTLRMPGQL IGDEGIHFIE AKRLISRRP
201 NDKEMKDFYD TIYIKRKEI EEEIKECFCR AL

!!A-SEQUENCE 1.0
ID YOTC_CAEEL STANDARD: PRT: 161 AA.
AC Q10120:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE HYPOTHETICAL 18.2 KD PROTEIN ZK632.13 IN CHROMOSOME III.
GN ZK632.13.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitia;
OC Rhabdilitia; Rhabdilitidae; Peloderinae; Caenorhabditis.
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRATON L., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULSTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA LATEILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA STULTON J., THIERRY-MIEG J., THOMAS K., THOMAS M., VAUGHAN K.,
RA WATKINSON P., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOLFDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";

```

```

Nature 368:32-38(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z22181; CAA80198.1;
DR WORMPEP; ZK632.13; CE01711.
KW Hypothetical protein.
SQ SEQUENCE 161 AA; 18180 MW; 9298BA33 CRC32;
YOTC_CAEEL Length: 161 February 14, 2000 08:03 Type: P Check: 9212

1 MSRLGFIQY EFGDEMFQ QMEIKSNAE QAKMEQCK MLEETEMPE
51 ESEVPVAKCL DPEAFQSES VSKGYESPK NISFLKEDAV TVNMSCPA
101 DDIAKLIRNI QNSYTLIGE EARGCRGRL LNVLPFGSA SPRLOPTTP
151 KNVALETTGS Q

!!A-SEQUENCE 1.0
ID YOXD_BACSU STANDARD: PRT: 238 AA.
AC P14803:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN RFP-PELB INTERGENIC REGION (EC 1.-.-)
DE (ORF238).
GN YOXD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-168;
RX MEDLINE: 91192601.
RA ANN K.S., WAKE R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes."
RT Gene 98:107-112(1991).
RL [2]
RP SEQUENCE OF 62-238 FROM N.A.
RC STRAIN-168;
RX MEDLINE: 88040469.
RA CARRIGAN C.M., HARRSWA J.A., SMITH M.T., WAKE R.G.;
RT "Sequence features of the replication terminus of the Bacillus
RT subtilis chromosome."
RT Nucleic Acids Res. 15:8501-8509(1987).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X06168; CAA29533.1;
DR EMBL: Z99114; CAB13743.1;
DR PIR: S01270; S01270.
DR HSSP: P19992; 2HS0.
DR SUBTILIST; BG11048; YOXD.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
KW Hypothetical protein; Oxidoreductase.
NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).

```


FT ACT_SITE 155 155 BY SIMILARITY
 SQ SEQUENCE 238 AA: 25299 MW: 06381861 CRC32:
 YPOG_BACSV Length: 238 February 14, 2000 08:03 Type: P Check: 2171 ..
 1 MOSLOHKTAL ITGGGGRIGR ATALALAKE VNIIGLIGRTS ANVEKVAEEV
 51 KALGVKAAPA AADVKADADOV NOAVQVKEQ IGDIDILINN AGISKEGFL
 101 DLSADEMENI IOVNLGVYH VTRAVLPENI ERKAGDINI SSTAGORGA
 151 VTSVYSASKF AVLGLESIM QEVKKNIRV SALTPTVAS DMSIELULTD
 201 GNEPKVMOPE DLAEYMAOL KIDPRIFIKT AGLMSTNP
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 285 AA.
 ID YPOG_ECOLI
 AC 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 15-DEC-1998 (rel. 39, Last annotation update)
 DE HYPOTHEICAL TRANSCRIPTIONAL REGULATOR IN DDG-GLK INTERGENIC REGION.
 GN YPOG.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 NC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 9742617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE: 97349980.
 RA YAMAGOTO T., AIBA H., BABA T., HAYASHI K., INADA T., ISONO K.,
 RA ITOH T., KIMURA S., KITAGAWA M., MAKINO K., MITI T., MITSURASHI N.,
 RA MIZOBUCHI K., MORI H., NAKADE S., NAKAMURA I., NASHIMOTO H.,
 RA OSHIMA T., OYAMA S., SATO N., SAMPEI G., SATOH Y., SIVASUNDARAM S.,
 RA TAGAMI H., TAKAHASHI H., TAKEEDA J., TAKEKOTO K., UEHARA K., WADA C.,
 RA YAMAGATA S., HORICHI T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -1- SIMILARITY: BELONGS TO THE ARAC/YLIS FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000326; AAC75441.1; -;
 CC EMBL: D90868; CAB22182.1; -;
 CC DR ECGENE; EGI4150; YPOG.
 CC DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
 CC DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
 CC DR PFAM; PF00155; HTH_2; 1.
 CC DR HYPOTHEICAL PROTEIN; Transcription regulation; DNA-binding.
 CC FT DNA_BIND 200 219 H-T-H MOTIF (BY SIMILARITY).
 CC SQ SEQUENCE 285 AA: 32355 MW: 61FPA84 CRC32:
 YPOG_ECOLI Length: 285 February 14, 2000 08:03 Type: P Check: 4986 ..

1 MKARGLPADQ QRFADLFSGI VLNPLLGRV WFAQSAPASLP VESLCIDFPR
 51 LDIYRGEGY NLEAKOQRL VEGEMLFIPA RAANLPVNNK PVMLSIVFA
 101 PTWGLSTFYD SRTSLIHPA ROIQLPSPOR GECEMITAL TLUSSPLEQ
 151 NIIOPLVSL LHCRSVYNN PPGNSOPROD ELYHSIONWY ONVYKOPLTR
 201 ESWAQFNIT PNHLSKLEAQ HGTMRFEIYV RWMAMAKAM ILOKYLHSH
 251 EWAQRGFPD SDYFCRVFRR OFGLTPGEYS ARFQG
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 224 AA.
 ID YPOG_BACSV
 AC 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEICAL 24.8 KD PROTEIN IN DAPB-PAPS INTERGENIC REGION.
 GN YPOG OR JOUG.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 NC Bacillus; Firmicutes; Bacillus/Clostridium group;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / MAREBURG;
 RX MEDLINE: 96349105.
 RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,
 RA SERROR P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kds loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 CC -1- SIMILARITY: TO E.COLI YAIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L38424; AAA92876.1; -;
 CC DR EMBL: L47709; AAB38444.1; -;
 CC DR EMBL: 289115; CAB14163.1; -;
 CC DR SUBTILIST; BG11212; YPOG.
 CC KW HYPOTHEICAL PROTEIN.
 CC SQ SEQUENCE 224 AA: 24806 MW: DF897440 CRC32:
 YPOG_BACSV Length: 224 February 14, 2000 08:03 Type: P Check: 3316 ..
 1 MYNADVLAFG AHSDDVEIGM GGTIAKVRQ EKKYMICDLT EAEIISNGTV
 51 SLKEKBAEA ARILGADKRI QLTLPDRGLI MSDAIRSIV TVIRICPRNA
 101 VEMPYKDRKH PDHGNAAALV EEAIFSAIGH KYKDEKSLPA HKVSKYVYVY
 151 INGFHQPDVY IDISDTIEAK KQSLNAYKSO FIPKDSVST PLTNGYIEIV
 201 EAREKLYGKE AGVEYAEVSF PNCC
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 271 AA.
 ID YPOG_KLEPN
 AC P27569;
 DT 01-AUG-1992 (rel. 23, Created)
 DT 01-AUG-1992 (rel. 23, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHEICAL PROTEIN IN POOA 5' REGION (ORF X) (FRAGMENT).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 NC Klebsiella.
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN-NCCTC 418:
RX MEDLINE: 92212293.
RA MEDLEBERG J.U.M., SELLINK E., RIEGMAN N.H., POSTMA P.W.:
RT Nucleotide sequence and structure of the Klebsiella pneumoniae ppg
operon.
RL Mol. Genet. 232:284-294(1992).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M19.
CC -1- SIMILARITY: TO AN ORF IN THE 3 REGION OF PQO-III IN
A.CALCOCOCTICUS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: X58778: CAA41578.1:
DR PIR: S20452: S20452.
DR PROSITE: PS00869: RENAL_DIPEPTIDASE; 1.
DR Hypothetical protein.
FT NON_TER
FT SEQUENCE 271 AA: 29484 MW: 135841F CRC32:
YPOQ_KLEPN Length: 271 February 14, 2000 08:03 Type: P Check: 9681 ..

1 SVPEDPIDIL WOQALILKOL IAHSGRLRL CLSAADIENC REDKVLANYA
51 HIEGAGEFDG EGRDLOAFYA AGVNSIGPFW NIANREGSGV NGSEFGSDPT
101 GGLTAAAGID LKQVNAALK QIDVSHNEK AFWDTAHHAT SPLVATHSNA
151 HALCPORNL TDQOLRAIRD SGGVGVNFG NAFRLADGR DSDPLTTIV
201 RHIDYLINIM GEDHVALGSD FDGITLPDEL GDVAGLPRLI NTLRAGYDQ
251 LVLDKLMRN WLRVKNVQ Q
11AA SEQUENCE 1.0
ID YPOE_ECOLI STANDARD: PRT: 425 AA.
AC P77031: P76634; P76635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEMETICAL 46.8 KD PROTEIN IN CYSJ-ENO INTERGENIC REGION.
GN YPOE
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.:
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12.
RX ATBA H., BABI T., FUJITA K., HAYASHI K., HONTO A., HORIUCHI T.,
RA IKEMOTO K., INABA T., ISONO K., ISONO S., ITOH T., KANA K.,
RA KASAI H., KASHIMOTO T., KIM S., KIMURA S., KITAGAWA M.,
RA KITAKAWA M., MARINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,
RA MOTOMURA K., NAKAMURA F., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RA SATO N., SAEPEI G., SEKI Y., TAGAWA H., TAKEMOTO K., WADA C.,
RA YAMAMOTO Y., YANO M.:
RT Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
```

```

CC -----
CC -1- SIMILARITY: TO E.COLI Y1HN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AE00361: AAC75817.1:
DR EMBL: D90893: CAE22504.1:
DR ECOGENE: E613174: YPOE.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 9
FT TRANSMEM 29
FT TRANSMEM 49
FT TRANSMEM 69
FT TRANSMEM 76
FT TRANSMEM 96
FT TRANSMEM 97
FT TRANSMEM 117
FT TRANSMEM 139
FT TRANSMEM 159
FT TRANSMEM 172
FT TRANSMEM 192
FT TRANSMEM 220
FT TRANSMEM 240
FT TRANSMEM 260
FT TRANSMEM 280
FT TRANSMEM 292
FT TRANSMEM 312
FT TRANSMEM 314
FT TRANSMEM 334
FT TRANSMEM 355
FT TRANSMEM 375
FT TRANSMEM 389
FT SEQUENCE 425 AA: 46830 MW: 886A1E5D CRC32:
YPOE_ECOLI Length: 425 February 14, 2000 08:03 Type: P Check: 1867 ..

1 MOHNSYRMI TLAIISFSG VSFDLAYLRY IYQIPMAKFM GFENTIEGLI
51 MSTFGIAIIT LYAPSGYAD KESHRMITS AMITGLLGL LMAITYPLWV
101 MLCIQAFPAI TTILMNSVS IKASLSDGH SEQCKIMGM EGRGVGWS
151 LAVFTWVES RPAPDSTSL KVIYIISVY YILGILCWF FVSDDNNLNS
201 ANNEKOSFO LSDILAVLRI STWYCSNVI FGVFTYAIL SYSTNYLTEN
251 YGMSLVASY MGIVINKIFR ALGCPGLGII TTSKVSPT RVLIQLSVLG
301 LITLITLAVT NSNPQVAMG IGLILGLFT CYASNGLYWA CPGEARTPSY
351 IMGTYGICS VIGFLPDVYV YPIIGHMODT LPAAEVYENM WLMGMALGM
401 VIVFTFLFO KIRTAADAPA MASSK
11AA SEQUENCE 1.0
ID YPOE_ECOLI STANDARD: PRT: 141 AA.
AC P77136:
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEMETICAL 16.3 KD PROTEIN IN KDU1-LISS INTERGENIC REGION.
GN YPOE
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.:
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1A-141 FROM N.A.
RC STRAIN-K12 / MG1655;
RX ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
```

RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
 LEW H., LIN D., NAMATH A., OEFNER P., SCHRAM S., DAVIS R.W.;
 Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000369; AAC75888.1; ALT_INIT.
 DR EMBL: U83187; AAB40288.1; -.
 DR ECGENE: E613274; yjek.
 DR Hypothetical protein.
 KW SEQUENCE 141 AA; 16271 MW; E96C8D18 CRC32;
 SQ
 YOEK_ECOLI Length: 141 February 14, 2000 08:03 Type: P Check: 8893 ..
 1 MDIEFSQIHE MYMHDIIVNS DSKKRPRL KFLNAENVL TQTSWTLS
 51 RYVNVSVNK VNKSKYKNS YISRYNDEF SLTDEINSE KETLVLSID
 101 SLSKVLNMP LSVLTSTVR RNNRKNV EFDWICTRC C
 11AA_SEQUENCE 1.0
 ID YR7B_ECOLI STANDARD; PRT; 169 AA.
 AC P21316;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE HYPOTHEICAL 18.1 KD PROTEIN (ORF) (RETRON EC67).
 DE Escherichia coli.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CL-1:
 RX MEDLINE: 91067724.
 RA HSU M.-Y., INOUE M., INOUE S.;
 RT "Retron for the 67-base multicopy single-stranded DNA from
 RT Escherichia coli: a potential transposable element encoding both
 RT reverse transcriptase and Dam methylase functions.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9454-9458(1990).
 CC -1- SIMILARITY: 70% IDENTITY TO CP76 OF BACTERIOPHAGE 186.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M55249; AAA23393.1; -.
 DR PIR: J00857; J00857.
 KW Transposable element; Hypothetical protein.
 SQ SEQUENCE 169 AA; 18129 MW; 9D36865A CRC32;
 YR7B_ECOLI Length: 169 February 14, 2000 08:03 Type: P Check: 6140 ..
 1 MEDYVSKHP HFDACRARA LKHNVQLAE RAGMNVQLR NKNPQPHL
 51 LTAPIWLLT DLTEDSLVD GFLAQHCLP CVPINEVAKE KLPHYMSAT
 101 AEIGVNAAGA VSGDVKTSAG RDAISSINS VTRLMALAV SLQARQANP
 151 AMAASVDIVT GLGASFGLL
 11AA_SEQUENCE 1.0
 ID YRBC_HAEIN STANDARD; PRT; 214 AA.
 AC P45028;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHEICAL PROTEIN H1084 PRECURSOR.
 GN H1084.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20.
 RX MEDLINE: 95350650.
 RA FIEDSCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BOLT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCKYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., RANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORHAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: STRONG, TO E.COLI YRBC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U33788; AAC2740.1; -.
 DR TIGR: H11084; -.
 KW Hypothetical protein; Signal.
 FT SIGNAL 28
 FT CHARIN 29 214
 SQ SEQUENCE 214 AA; 24510 MW; 02256B8A CRC32;
 YRBC_HAEIN Length: 214 February 14, 2000 08:03 Type: P Check: 1216 ..
 1 MNLQLKMF TLTFVLNAF LVTRTAIAET SPVYLMOQAA DLFSDIOAN
 51 QSKIKODPNY LRTIVNDLL PYVNLVYAS KVLGSYYKST SAOREKEFFK
 101 TFGELIECKY AQALVNSNQ KIOIESEKEL GDNNFNIRY NIIQANGVAP
 151 ILLFKVRKG NKSGEKVID MVGAGVSMLE DTRKNVGIL NKGIDTLIT
 201 KMOQASAPI IFNO
 11AA_SEQUENCE 1.0
 ID YRKG_BACSU STANDARD; PRT; 36 AA.
 AC P54434;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOTHEICAL 4.2 KD PROTEIN IN BLTR-SPOIIC INTERGENIC REGION.
 GN YRKG.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-168 / JH642.
 RA KOBAVASHI Y., MIZUNO M., MASUDA S., TAKEKAWA K., HOSONO S.,
 RA SAO T., TAKEUCHI M.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D84432; BAA1362.1; -
DR EMBL: Z99117; CAB14593.1; -
DR SUBMITTER: BG11773; YKRG.
KW Hypothetical protein.
SQ SEQUENCE 36 AA; 4234 MW; C3BF5DEA CRC32;

YKRG_BACSU Length: 36 February 14, 2000 08:03 Type: P Check: 1724 ..

1 MYKAMPRE VIKKSLTKNL YLFMVCYMKM ILVIYR

!!AA_SEQUENCE 1.0
ID YK73_MYCTU STANDARD; PRT; 118 AA.
AC P85085;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 13.8 KD PROTEIN RV3073C.
GN RV3073C OR MTCY22D7.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE: 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLEMEIER K., GAS S., BARRY C.E. III, TERAIA F.,
RA DAVOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA BAYES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMILTON R.,
RA HORNSBY T., JAGELS K., KROCH A., MCLELLAN J., MOLE S., MORPH L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RABANDRAN M.A., ROGERS J.,
RA RUTHER S., SEEGER K., SKELTON S., SOKARS S., SOKARS R., STOLSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RT Nature 393:527-544(1998).
RL 1. SIMILARITY: TO E.COLI YKAO.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z83866; CAB06250.1; -
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 13750 MW; B38D283F CRC32;

YU73_MYCTU Length: 118 February 14, 2000 08:03 Type: P Check: 3649 ..

1 MYRERVRVA RYVEDIDPD GORVLDRIIM PHGIRKDOOR VGIMCKRYAP

51 SKLEWYHH OPEFDEFFAS ROGEELHDSA ALAEIRKLIG RSVYTPVAT
101 RHVAASHAAV LAOLLNR

!!AA_SEQUENCE 1.0
ID YK73_MYCTU STANDARD; PRT; 394 AA.
AC P38642;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHEICAL 43.4 KD PROTEIN IN ROCC-PTA INTERGENIC REGION.
KW YWF OR IPA-84D.
RC Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE: 95020337.
RA GLASER P., KUNST F., ARNOLD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONSCU M., LUBCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT Mb region from 325 degrees to 333 degrees."
RL Mol. Microbiol. 10:371-384(1993).
CC -1. SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X73124; CA51640.1; -
DR EMBL: Z98123; CAB15797.1; -
DR PIR: S39739; S39738.
DR SUBMITTER: BG10630; YWFF.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11
FT TRANSMEM 31
FT TRANSMEM 43
FT TRANSMEM 63
FT TRANSMEM 69
FT TRANSMEM 92
FT TRANSMEM 92
FT TRANSMEM 142
FT TRANSMEM 142
FT TRANSMEM 166
FT TRANSMEM 166
FT TRANSMEM 215
FT TRANSMEM 215
FT TRANSMEM 244
FT TRANSMEM 244
FT TRANSMEM 288
FT TRANSMEM 308
FT TRANSMEM 332
FT TRANSMEM 332
FT TRANSMEM 353
FT TRANSMEM 373
SQ SEQUENCE 394 AA; 43416 MW; 93D522A2 CRC32;

YWFF_BACSU Length: 394 February 14, 2000 08:03 Type: P Check: 3644 ..

1 MKOLKNSKY LLYGALSFM GDYCVLPALL ILSTYHDW VTSQVIYRS

51 IPVWFQPIFG VLVDRDLRIK IMLTDLIRG IIFLGITFLP KGEYPLITLA
101 LLEITGSGV FENPARLAVM SLESDIKSI NTLPAKATTI SIYGAAGG
151 LFLGGSVEL AVAFNGVYLL VSAFFISRIK LQFVPIQSEN IKAEPQSFKE
201 GLKEIKTNSF VLNAMFTMIT MALMGVYS YPIVSRFLG DGEIGNFILT
251 FCGFSGFTG AALVSKMGFN NNRGLTYFTV LSIVALFL FTPIFAVSI
301 AALIFIANE YGEVIAKYV QENANQIOG RIFSVAESI GLCISISGMF
351 INILSAVIM GLIIVYVGL FLHTKLVRK FLERDKTEQ KGVF

!!AA_SEQUENCE 1.0
ID YK73_MYCTU STANDARD; PRT; 274 AA.
AC Q93834;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 31.6 KD PROTEIN F596.8 IN CHROMOSOME I.
GN F596.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;

RA WILKINSON J.:
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO A FAMILY OF C.ELEGANS PROTEINS THAT GROUPS
 CC C37A7.2, C33H5.2, C35A5.5, F13G3.3, F59C6.8, F49C12.5, R07B7.12
 CC AND ZK381.2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z79600; CAB01879.1; -
 CC DR WOMMPP; F59C6.8; C31470.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 274 AA; 31624 MW; 43119A02 CRC32;
 YYVP_CAEEL Length: 274 February 14, 2000 08:03 Type: P Check: 1176 ..
 1 MGNVEVOFK IOPVINTIPE FCKWVPYLAV GOVEDHVLV KLSTNKIDGM
 51 ELSRTPEYET PRKVACFSP LELNERWOLL LATVEIYSHY GAFMHFYTRI
 101 SPMSAIRIGE SRAASPMFDP NTELEFRNOA SAMTDLLOY KEAEFIVFP
 151 DPDDLVPVL GKNYEEFTQ AFKEPTAGA VYVNTQTSI ESSMTPALYS
 201 PISMLASMKF KGEQKMGKLV VPERVDSTW IHRSAIKES FQKVMFVDV
 251 NAFYLRIRWK FPEVPTFNRS SELL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 364 AA.
 ID ZRP4_MAIZE
 AC P47917;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE O-METHYLTRANSFERASE ZRP4 (EC 2.1.1.-) (OMT).
 GN ZRP4.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Zea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, NK31; TISSUE=ROOT;
 RA MEDLINE: 94105316.
 RX HEID B.M., WANG H., JOHN I., WURTELE E.S., COLBERT J.T.;
 RT "An mRNA putatively coding for an O-methyltransferase accumulates
 RT preferentially in maize roots and is located predominantly in the
 RT region of the endodermis."
 RT Plant Physiol. 102:1001-1008(1993).
 RL
 CC -1- FUNCTION: MAY BE INVOLVED IN THE O-METHYLATION OF SUBERIN
 CC PHENYLPROPANOID PRECURSORS.
 CC -1- TISSUE SPECIFICITY: ACCUMULATES PREFERENTIALLY IN THE ROOTS AND IS
 CC LOCATED PREDOMINANTLY IN THE REGION OF THE ENDODERMIS, LOW LEVELS
 CC ARE SEEN IN THE LEAVES, STEMS, AND OTHER SHOOT ORGANS.
 CC -1- SIMILARITY: TO OTHER OMTS REQUIRING S-ADENOSYL-L-METHIONINE AS
 CC SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L14063; AAA18532.1; -
 CC DR MAIZEDB; 63528; -
 CC PFAM: PF00891; Methyltransf_2; 1.

KW Transferase; Methyltransferase.
 SQ SEQUENCE 364 AA; 39583 MW; 3067DB07 CRC32;
 ZRP4_MAIZE Length: 364 February 14, 2000 08:03 Type: P Check: 6645 ..
 1 MELSPNNSTD QSLDAQLEL WHTTFAPMKS MALKSAHLR IADAIHNGG
 51 AASLSQILSK VLAHSRYSS LRRLMVLIT TVEFGTOLG GGSDDSEPV
 101 YTLTPYSRLI IGSOSQIAQ TPLAAVLDP TIVSPFSELG AMFOHELPPD
 151 CTRKHTGRG IWELTKDPT FDLAVNDGIA SDSQILVDVA IKOSAEPVQG
 201 ISSLVVGGG IGAAQAISK APFVKCSVL DLAHVAKAP THTDVQFIAG
 251 DMFESIPPAD AVLLKSVLHD WDHDCVKIL KNCKRAIPPR EAQGVIIIN
 301 MVGAGPSDM KHKMQAIFD VYIMFNGME RDEDEWSKIF SEAGYSDYRI
 351 IPLGVRSII EYYP

